From: Chan, Christina
Sent: Monday, June 20, 2005 8:55 AM
To: Yu, Misook; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 10/030,294

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----
From: Yu, Misook
Sent: Monday, June 20, 2005 5:48 AM
To: Chan, Christina
Subject: Rush search request for 10/030,294

Pls approv rush search. It is due this biweek.

Stic,
pls search SEQ ID NO:1.

Examiner Misook Yu, Ph.D.
571-272-0839 (Phone)
Art Unit 1642
REM-3A18 (Room)
REM-3C18 (Mail Box)

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search

NA#:_____ AA#:_____
Interference:_____ SPDI:___
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure#:_____ Text:___
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

THIS PAGE BLANK (USPTO)

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June 22, 2005, 14:15:40 ; Search time 23453 Seconds
                                         (without alignments)
                                         11726.935 Million cell updates/sec

Title:          US-10-030-294-1
Perfect score:  5676
Sequence:       1 cagctgttcagggatgtctg........tttatataattgttctccgt 5676

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters:      9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
             1: gb_ba:*
             2: gb_htg:*
             3: gb_in:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pl:*
             9: gb_pr:*
            10: gb_ro:*
            11: gb_sts:*
            12: gb_sy:*
            13: gb_un:*
            14: gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	5676	100.0	5676	6	BD093089	BD093089 Gene enco
	2	5676	100.0	5676	6	BD093091	BD093091 Gene enco
	3	5676	100.0	5676	6	BD093092	BD093092 Gene enco
	4	5676	100.0	5960	6	BD093090	BD093090 Gene enco
	5	5673.4	100.0	13940	6	BD093100	BD093100 Gene enco
c	6	5635.6	99.3	162646	9	AC078809	AC078809 Homo sapi
c	7	669.4	11.8	160574	2	AC135489	AC135489 Rattus no
c	8	669.4	11.8	284005	2	AC113784	AC113784 Rattus no
c	9	662	11.7	204313	2	AC126304	AC126304 Rattus no
c	10	649	11.4	214853	10	AC113020	AC113020 Mus muscu
c	11	372.6	6.6	56268	2	AC091240	AC091240 Mus muscu
	12	215	3.8	835	9	HSP63G01	AF124528 Homo sapi
c	13	212.4	3.7	159249	9	AC103882	AC103882 Homo sapi
c	14	205	3.6	60323	2	AC087334	AC087334 Homo sapi
	15	204.4	3.6	173456	9	AL158198	AL158198 Human DNA
c	16	201.2	3.5	150846	2	AC036173	AC036173 Homo sapi
c	17	200.6	3.5	170837	2	AC135613	AC135613 Pan trogl
c	18	200	3.5	195808	2	AL591182	AL591182 Homo sapi
c	19	196.6	3.5	125377	2	AC118135	AC118135 Homo sapi
c	20	196.6	3.5	181442	AC134875	AC134875 2	Homo sapi
c	21	196.6	3.5	202827	AC016907	AC016907 9	Homo sapi
	22	196.2	3.5	219122	AC080038	AC080038 9	Homo sapi
c	23	196	3.5	148911	AL353142	AL353142 2	Homo sapi
	24	196	3.4	273403	AC011498	AC011498 9	Homo sapi
	25	195	3.4	47028	AC117488	AC117488 2	Homo sapi
	26	194	3.4	163246	AC068401	AC068401 2	Homo sapi
c	27	194	3.4	174428	AC005821	AC005821 9	Homo sapi
c	28	193.4	3.4	118384	AF389951	AF389951 9	Homo sapi
	29	193.2	3.4	196467	AC022392	AC022392 9	Homo sapi
	30	193	3.4	139218	AC019289	AC019289 9	Homo sapi
	31	193	3.4	151880	AP005716	AP005716 9	Homo sapi
	32	193	3.4	166093	AC106013	AC106013 9	Homo sapi
	33	193	3.4	175931	AC026618	AC026618 9	Homo sapi
	34	193	3.4	180776	AC032043	AC032043 9	Homo sapi
	35	193	3.4	183336	AC108688	AC108688 9	Homo sapi
	36	193	3.4	291641	AC092893	AC092893 2	Homo sapi
c	37	191.8	3.4	179305	AC144411	AC144411 2	Homo sapi
	38	191.2	3.4	131757	AC004540	AC004540 9	Homo sapi
c	39	190.6	3.4	79340	AL359473	AL359473 2	Human DNA
c	40	190.6	3.4	161416	AC026857	AC026857 9	Homo sapi
	41	190.6	3.4	218074	HUAC002044	AC002044 9	Human Chr
c	42	190.4	3.4	157136	AC068251	AC068251 2	Homo sapi
	43	190.4	3.4	165240	AP002795	AP002795 9	Homo sapi
c	44	190.4	3.4	165878	AC073382	AC073382 2	Homo sapi
c	45	190.4	3.4	175006	AC027589	AC027589 2	Homo sapi

ALIGNMENTS

```
RESULT 1
BD093089

LOCUS       BD093089               5676 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Gene encoding promoter region of tumor suppressor gene p51.
ACCESSION   BD093089
VERSION     BD093089.1  GI:22638677
KEYWORDS    WO 0100818-A/1.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5676)
AUTHORS     Sakai,T., Kagaya,S., Sato,T., Sukenaga,Y. and Fujii,H.
TITLE       Gene encoding promoter region of tumor suppressor gene p51
JOURNAL     Patent: WO 0100818-A 1 04-JAN-2001;
            TOSHIYUKI SAKAI,NIPPON KAYAKU KK,SHIGEHIDE KAGAYA,TAKAMICHI SATO,
            YOSHIKAZU SUKENAGA,HIDEJI FUJII
OS          Homo sapiens (human)
PN          WO 0100818-A/1
PD          04-JAN-2001
PF          28-JUN-2000 WO 2000JP004261
PR          29-JUN-1999 JP 99P 183195
PI          TOSHIYUKI SAKAI,SHIGEHIDE KAGAYA,TAKAMICHI SATO,YOSHIKAZU PI
            SUKENAGA,
            PI   HIDEJI FUJII
            PC   C12N15/10,C12N5/10,C12Q1/68,A61K48/00,A61K45/00,A61P43/00, PC
            A61P35/00
CC
FH          Key             Location/Qualifiers
FEATURES             Location/Qualifiers
            source   1..5676
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             100.0%;  Score 5676;  DB 6;  Length 5676;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 5676;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 CAGCTGTTCAGGGATGTCTGAAAAGAAGAGCCACCCACATTGCTTCTGGACACTGGGTGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1 CAGCTGTTCAGGGATGTCTGAAAGAAGCCACCACATGCTCTGGACACTGGGTGT                60

Db   61 GACTTTGGAGGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCAAT         120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 GACTTTGGAGGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCAAT         120

Db  121 TGGCCCTGTCCCTTGCATGCCCTCTTCCTTCCCTTGGGACACTCCTTCCTTGGGACACTCCTTCCTTG   180
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy  121 TGGCCCTGTCCCTTGCATGCCCTCTTCCTTCCCTTGGGACACTCCTTCCTTG                180

Db  181 ACATTAACTTAACTATAAATGTTTATTGATGAATTCAGTGACCTGAAGAGAGATGGAG           240
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181 ACATTAACTTAACTATAAATGTTTATTGATGAATTCAGTGACCTGAAGAGAGATGGAG           240

Db  241 GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTGCTTTTTCATTGAATTAA         300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTGCTTTTTCATTGAATTAA         300

Db  301 AGTCATTCGAATACCATTCAGTTTACTTAGGTCTCAGGCCCAGCTTTACTTCATCCTAATCGAT      360
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 AGTCATTCGAATACCATTCAGTTTACTTAGGTCTCAGGCCCAGCTTTACTTCATCCTAATCGAT      360

Db  361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCCATGA       420
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCCATGA       420

Db  421 CAGAACTTTGGCTTCCACTTTACTAAAAATCACCAAACTCAGTCTCTTGAGCAATAGT         480
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  421 CAGAACTTTGGCTTCCACTTTACTAAAAATCACCAAACTCAGTCTCTTGAGCAATAGT         480

Db  481 CAGAAAAGTGTCATCAATTACCAAAATCAGTCTCTTCAATTGAGCAATAGT              540
          ||||||||||||||||||||||||||||||||||||||||||||
Qy  481 CAGAAAAGTGTCATCAATTACCAAAATCAGTCTCTTCAATTGAGCAATAGT              540

Db  541 TGGTGAATTACTCCACCACCTCCTCCTGAAGGTTCTTCCTGCTCCTCCTCCACTATA        600
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  541 TGGTGAATTACTCCACCACCTCCTCCTGAAGGTTCTTCCTGCTCCTCCTCCACTATA        600

Db  601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCGTTACCCTGACACAAGGAAT    660
          |||||||||||||||||||||||||||||||||||||||||||||
Qy  601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCGTTACCCTGACACAAGGAAT    660

Db  661 TCAGTTCTCTGATCTCATAGTCACAGGCTGCACAGAGCTCGCAGAGCTCTACGGCAAGATCAT       720
          |||||||||||||||||||||||||||||||||||||||||||||
Qy  661 TCAGTTCTCTGATCTCATAGTCACAGGCTGCACAGAGCTCGCAGAGCTCTACGGCAAGATCAT       720

Db  721 CTGCTTTAAGCCTCCTGTGGTGGCATCCGTGTCTTCCACTGCCCTGTACCTATTGCTCT          780
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy  721 CTGCTTTAAGCCTCCTGTGGTGGCATCCGTGTCTTCCACTGCCCTGTACCTATTGCTCT          780

Db  781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCGAAAACTCTCTGCTCTCGAAAACTCT   840
          |||||||||||||||||||||||||||||||||||||||||||||
Qy  781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTTGAAAACTCTCGAAAACTCTCTGCTCTCGAAAACTCT   840

Db  841 CCATCAGTCCACATGATCAGGCCTCTCCTGCCCAAACATGGCCCAAACATGGGAATT           900
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  841 CCATCAGTCCACATGATCAGGCCTCTCCTGCCCAAACATGGCCCAAACATGGGAATT           900

Db  901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAATTCCTATTCCCTGAACTTATAATTTAT          960
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy  901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAATTCCTATTCCCTGAACTTATAATTTAT          960

Db  961 AGTGGCTTGCAAAGTGACTGTCCATTCCTATTTCTTAAGTCCCTAGTCATCATGGTAGGG        1020
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  961 AGTGGCTTGCAAAGTGACTGTCCATTCCTATTTCTTAAGTCCCTAGTCATCATGGTAGGG        1020

Db 1021 CCTGGTTACAGCCCCTTTCTGAGATGTGTGTTTTTTTCCAACTGTCTCCTTATAGTCTGT       1080
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1021 CCTGGTTACAGCCCCTTTCTGAGATGTGTGTTTTTTTCCAACTGTCTCCTTATAGTCTGT       1080

Db 1081 GAATTTCATATTTCTTTTCATACATTTCATGTTTGTTTTGTGTTGTTTTTT                 1140
          |||||||||||||||||||||||||||||||||||||||
Qy 1081 GAATTTCATATTTCTTTTCATACATTTCATGTTTGTTTTGTGTTGTTTTTT                 1140
```

```
Db 1141 GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTATACCCAAGGAATCCTGATTGATACA        1200
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1141 GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTATACCCAAGGAATCCTGATTGATACA        1200

Db 1201 TCCTTCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGG         1260
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1201 TCCTTCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGG         1260

Db 1261 TCTGGGAGTAAGTTAGTAGTACCAGAGCTCGTACTACTAACCCCCAGGTTAGCCAACTGCTTTACAC   1320
          |||||||||||||||||||||||||||||||||||||||||||||
Qy 1261 TCTGGGAGTAAGTTAGTAGTACCAGAGCTCGTACTACTAACCCCCAGGTTAGCCAACTGCTTTACAC   1320

Db 1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAGAAGAGCTACTATTTTGC        1380
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAGAAGAGCTACTATTTTGC        1380

Db 1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATAATGTGATGATTGAACTCTTAAAA       1440
          |||||||||||||||||||||||||||||||||||||||||||||
Qy 1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATAATGTGATGATTGAACTCTTAAAA       1440

Db 1441 AGTTTGTTGGACTTCTGCGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT        1500
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1441 AGTTTGTTGGACTTCTGCGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT        1500

Db 1501 AAACAAACAACAAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC         1560
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1501 AAACAAACAACAAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC         1560

Db 1561 AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC        1620
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1561 AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC        1620

Db 1621 ACAGAGAATACGTACCATATGCCTGCATTTATGTGATATTCTAGCATTGTATTGTTGTCCAA       1680
          |||||||||||||||||||||||||||||||||||||||||||||
Qy 1621 ACAGAGAATACGTACCATATGCCTGCATTTATGTGATATTCTAGCATTGTATTGTTGTCCAA       1680

Db 1681 CATAGTAGCCAGTGCCTCACATGGCTATTCAAATTTAAATTGATTGAATTAAGTAAGA         1740
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1681 CATAGTAGCCAGTGCCTCACATGGCTATTCAAATTTAAATTGATTGAATTAAGTAAGA         1740

Db 1741 ATAAAAATTTAGCTCTTCAGTAGCCTTAGCCACATCGAACATCAGTAACATCAGTAACATCAGAC     1800
          |||||||||||||||||||||||||||||||||||||||||||
Qy 1741 ATAAAAATTTAGCTCTTCAGTAGCCTTAGCCACATCGAACATCAGTAACATCAGTAACATCAGAC     1800

Db 1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG        1860
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG        1860

Db 1861 AAAATTACACATAATCTCATAAAACAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGGG        1920
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1861 AAAATTACACATAATCTCATAAAACAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGGG        1920

Db 1921 TGAGGGGAGATCGATTGCAAAGTGGTATGGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA        1980
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1921 TGAGGGGAGATCGATTGCAAAGTGGTATGGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA        1980

Db 1981 ATCTTGATTGCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT        2040
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1981 ATCTTGATTGCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT        2040

Db 2041 TGGAATTGGCGAATTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA         2100
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2041 TGGAATTGGCGAATTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA         2100

Db 2101 AGTGTTTTGTGTGGAAATATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA        2160
          |||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2101 AGTGTTTTGTGTGGAAATATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA        2160

Db 2161 AATACATAAACATGTATTCATCCATTAGGTGCAATATTTTTGCTAGCTCCTGAAAACAC        2220
          ||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2161 AATACATAAACATGTATTCATCCATTAGGTGCAATATTTTTGCTAGCTCCTGAAAACAC        2220
```

```
Qy  3301                                                              3360
Db  3301 CATGCCTGTAATCCCAACACTTTGGGAGGAAGCCAAGGCCGGGTCGGGGATGGATAGCTTGAGTCCAGCAAAACAA 3360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3361                                                              3420
Db  3361 TTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAACAAAACAAAAACAA 3420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3421                                                              3480
Db  3421 AATTACCTGGGTATGGTGGTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGA 3480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3481                                                              3540
Db  3481 AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACTC 3540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3541                                                              3600
Db  3541 AGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAATAATAAATAAAATAACTTTT 3600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3601                                                              3660
Db  3601 AAAAAACAAAAATTAATTAAATTTTAAAAACACAACACACACTAGAGATGTTGCAAATTGA 3660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3661                                                              3720
Db  3661 TTATTTGGGAGTCTATATCCCTGGAAGTTGTAAATATCAGAGCTAGAAGAGTTCTTCCTCA 3720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3721                                                              3780
Db  3721 TTTCCTAGAGCGTCGAATTGTAAATATCAGAGCTAGAAGAACACTTAGGGCTCGCCCACT 3780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3781                                                              3840
Db  3781 CCAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAG 3840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3841                                                              3900
Db  3841 AGTCTTAGGGCTTCACCCCAGAACCTGACCCAGAATCTGCATTAACAAGATTTCTAGGT 3900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3901                                                              3960
Db  3901 GCCTCACGGGCACATTAAAACTTGAGAAGTTCTGCACTAGAAAATCTTCACTCCACCTTTC 3960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  3961                                                              4020
Db  3961 ATTATAAATGAATCACTTGGCTGTGGTCACAGGAGAAATTGATTATTTTTAAATTCAGAA 4020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4021                                                              4080
Db  4021 CCTTCTATTTAGGTCATCTATATTGCTAATAGCAGGGAGACTAAGGGGCCAAACTCTTTAACT 4080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4081                                                              4140
Db  4081 GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG 4140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4141                                                              4200
Db  4141 TGGAGCAAGCTGTTTGATTTGGCTGCGGGGCTCAGGGCCTGTTCGTGAATTTCACAATT 4200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4201                                                              4260
Db  4201 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGAGAGAATGTCAAGTTTTAAATAGCT 4260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4261                                                              4320
Db  4261 TCTCCCTTCCATCCTGCTGAAGCAACAAAATAAAATATTTTATGAAACACATTTTGAGT 4320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4321                                                              4380
Db  4321 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTGAAGGGCTTTAGATTGTCTCACAAC 4380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  4381 TTTGACATCTACTGATGTCACCTATTTACAGTGTGTCCTGTGACTGGGGTGAAGGGA 4440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  2221 AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACGAGGAGATG 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACGAGGAGATG 2280

Qy  2281 AGTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAA 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 AGTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAA 2340

Qy  2341 GTGCAGATATAATTAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTATGTGAAAAA 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GTGCAGATATAATTAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTATGTGAAAAA 2400

Qy  2401 GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCCGCAGATGGAGTTT 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCCGCAGATGGAGTTT 2460

Qy  2461 TAAAACATTCCTGGTGGAGGGCAGATATGATCCAAGGCACAACCAGAGAAAATA 2520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 TAAAACATTCCTGGTGGAGGGCAGATATGATCCAAGGCACAACCAGAGAAAATA 2520

Qy  2521 TGCAACCTAGAGGAAGTGCATGAAGGGGAGCAGTGTTGGCGAGTGCAGTTGTTAAAATGTAA 2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 TGCAACCTAGAGGAAGTGCATGAAGGGGAGCAGTGTTGGCGAGTGCAGTTGTTAAAATGTAA 2580

Qy  2581 GTGAGAAGAATTTGTATCATAGACACCTGAGTTTTGCGAGGTGCATGTTCTTGGCTCCTA 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 GTGAGAAGAATTTGTATCATAGACACCTGAGTTTTGCGAGGTGCATGTTCTTGGCTCCTA 2640

Qy  2641 GGAGTCAAGAAGAACAAAGTGTCCCTTTCTTCTCCACGTTATGCTCAGTGGTCCAAGTCCAA 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 GGAGTCAAGAAGAACAAAGTGTCCCTTTCTTCTCCACGTTATGCTCAGTGGTCCAAGTCCAA 2700

Qy  2701 AACACCTTTCCTTCTTAAGTACTTCTTCTCCCCTCCATACAAATCTAAAGTCTTCACA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 AACACCTTTCCTTCTTAAGTACTTCTTCTCCCCTCCATACAAATCTAAAGTCTTCACA 2760

Qy  2761 AACATCATTTAAAACAGGCAGTCATGGTCAGAAAGCAATTGCTTTTGCTTTGCTTCTAT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 AACATCATTTAAAACAGGCAGTCATGGTCAGAAAGCAATTGCTTTTGCTTTGCTTCTAT 2820

Qy  2821 GTACGTTATTATATTACAATTTCTGCCTAAAAGACTCTAAAGTCTTGGAAAAGTTCCAC 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 GTACGTTATTATATTACAATTTCTGCCTAAAAGACTCTAAAGTCTTGGAAAAGTTCCAC 2880

Qy  2881 CTTCTTTTTCCTGGGTCAACTTGTATAGCATTGTATAGTAGTAACCTTAGTCCCCTAAGAGAAT 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTTCTTTTTCCTGGGTCAACTTGTATAGCATTGTATAGTAGTAACCTTAGTCCCCTAAGAGAAT 2940

Qy  2941 AAGGATGAACTATAAATATAAGAAGTAATTATGGTAATTATAATAATGATTGCCACTTATT 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 AAGGATGAACTATAAATATAAGAAGTAATTATGGTAATTATAATAATGATTGCCACTTATT 3000

Qy  3001 TTTCACTTGATCGTGTATGGTGTCATGCTACTGGTGTTCTGTTGAATTCTAGAGAGTTTG 3060
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001 TTTCACTTGATCGTGTATGGTGTCATGCTACTGGTGTTCTGTTGAATTCTAGAGAGTTTG 3060

Qy  3061 CCTCTTTTTCCTGGGTCAACTCCGCCATTTATTTCCGCCATTTATTTCCATATAATGCAATCTT 3120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061 CCTCTTTTTCCTGGGTCAACTCCGCCATTTATTTCCGCCATTTATTTCCATATAATGCAATCTT 3120

Qy  3121 TTTCATAATTACTTATTTAAAATTTGTTGCCATTTAAAATTTGTTGCTTAATAATGATATTATT 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121 TTTCATAATTACTTATTTAAAATTTGTTGCCATTTAAAATTTGTTGCTTAATAATGATATTATT 3180

Qy  3181 ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATT 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181 ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATT 3240

Qy  3241 TAAATACGTTAGGCTATAAACCTTTTAAAATTTTTTAAAAATAGATGAGTGGTGGCT 3300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241 TAAATACGTTAGGCTATAAACCTTTTAAAATTTTTTAAAAATAGATGAGTGGTGGCT 3300

Qy  3301 CATGCCTGTAATCCCAACACTTTGGGAGGAAGCCAAGGCCGGGTCGGGGACGGATAGCTTGAGTCCAGCAG 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  4381  TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGGA  4440

Qy  4441  AGATGTGAACTCACCATGTTAGTGACCGTTAGATCACAGAGTGTTTTTTTCCCCCTG  4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441  AGATGTGAACTCACCATGTTAGTGACCGTTAGATCACAGAGTGTTTTTTTCCCCCTG  4500

Qy  4501  TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCAAATCCACAAA  4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501  TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCAAATCCACAAA  4560

Qy  4561  ACCAGGATAAGTTTACAGCCCATATTCAGAAAGAAATAAATATTTGTGTGTAGACTT  4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561  ACCAGGATAAGTTTACAGCCCATATTCAGAAAGAAATAAATATTTGTGTGTAGACTT  4620

Qy  4621  TCCTGATATTACACTGATTTGGGAATATGAATAATGAACAATTTTATGGTTTCCTTTCCGAAGTAG  4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  TCCTGATATTACACTGATTTGGGAATATGAATAATGAACAATTTTATGGTTTCCTTTCCGAAGTAG  4680

Qy  4681  GTCAAGTCAAAGCAAAACCAAAAACAGCAAAAAACTGTAAGACATAAAGAGATGGAG  4740
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  4681  GTCAAGTCAAAGCAAAAGCAAAAACAGCAAAAAACTGTAAGACATAAAGAGATGGAG  4740

Qy  4741  CCGACTGAGAGATTAAAATAAAACTAGAAATATTTTTATTAACAGGCAATTTGAAATAATTT  4800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4741  CCGACTGAGAGATTAAAATAAAACTAGAAATATTTTTATTAACAGGCAATTTGAAATAATTT  4800

Qy  4801  GTGCACTTCAGAATATTCTCACAATATATATATATTATTTCCAATTTTAATATCTTTAAGAAAA  4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4801  GTGCACTTCAGAATATTCTCACAATATATATATATTATTTCCAATTTTAATATCTTTAAGAAAA  4860

Qy  4861  TTACTATATATATGTAAGTACATGTGCATGTGTTGTTGAGGTAGGATATATTAACTCAATAA  4920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4861  TTACTATATATATGTAAGTACATGTGCATGTGTTGTTGAGGTAGGATATATTAACTCAATAA  4920

Qy  4921  AGGTTATTTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCT  4980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4921  AGGTTATTTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCT  4980

Qy  4981  CTTTCTTAGCTGGAGGAAGAGTGAGTTCTAAGTAAATATAATCAAGGAATTTCCCT  5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  CTTTCTTAGCTGGAGGAAGAGTGAGTTCTAAGTAAATATAATCAAGGAATTTCCCT  5040

Qy  5041  GTCTTTGCTATTTGAGATTGTGACCACAACGGCGGTTGCCGTTGGAAACTGAAGGG  5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  GTCTTTGCTATTTGAGATTGTGACCACAACGGCGGTTGCCGTTGGAAACTGAAGGG  5100

Qy  5101  CGGGGAGGGAGGGAAATAGATGAAAAAACAAAACAAAAACAAAACTTCCCCAAGCAGCTCT  5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  CGGGGAGGGAGGGAAATAGATGAAAAAACAAAACAAAAACAAAACTTCCCCAAGCAGCTCT  5160

Qy  5161  ACAAAAACATTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAAACCAGTATCCAGA  5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5161  ACAAAAACATTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAAACCAGTATCCAGA  5220

Qy  5221  TACAAGGAAGTGTTATGTGAGCTGGAGCAGGGTGGACACTCAGTTCAGTTCAGTTACA  5280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221  TACAAGGAAGTGTTATGTGAGCTGGAGCAGGGTGGACACTCAGTTCAGTTCAGTTACA  5280

Qy  5281  AAAGTTCCAGGCTGCTGAAATTAAACTCTGAAATTAAACTCTGATGCCATCATGCCAGCATCCAATCACGACA  5340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281  AAAGTTCCAGGCTGCTGAAATTAAACTCTGAAATTAAACTCTGATGCCATCATGCCAGCATCCAATCACGACA  5340

Qy  5341  GAGATCAGAGTTCAGAGATGCCTCAGAGTGCCAACAACAGTGTGGCTACTA  5400
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341  GAGATCAGAGTTCAGAGATGCCTCAGAGTGCCAACAACAGTGTGGCTACTA  5400

Qy  5401  TACGTCAAGGACTCTGAAGCCGTGAGAGAGAAGTTTATGAAGTTTATGAAGTGTCAATTGATGAATCTCATTGGCTAAA  5460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401  TACGTCAAGGACTCTGAAGCCGTGAGAGAGAAGTTTATGAAGTTTATGAAGTGTCAATTGATGAATCTCATTGGCTAAA  5460

Qy  5461  GCTGGTAAGAATCGAGTGTGTTTATGAAGTTTATGAAGTTTGCAATGGCACTGCATTGGCTAAA  5520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461  GCTGGTAAGAATCGAGTGTGTTTATGAAGTTTATGAAGTTTGCAATGGCACTGCATTGGCTAAA  5377
```

Query Match 100.0%; Score 5676; DB 6; Length 5676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1  CAGCTGTTCAGGGATGCTCTGAAAAGAAGCCCACCCACCATTGCTTCTTCGGACACTGGGTGT    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5676  CAGCTGTTCAGGGATGCTCTGAAAAGAAGCCCACCCACCATTGCTTCTTCGGACACTGGGTGT  5617

Qy    61  GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCCTGCCCCAAT   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5616  GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCCTGCCCCAAT  5557

Qy   121  TGGCCCTCGTTCCCTGCCATGCCCTCTTTCCTGGACCATCTCTTG   180
          |||||||||||||||||||||||||||||||||||||||||||||
Db  5556  TGGCCCTCGTTCCCTGCCATGCCCTCTTTCCTGGACCATCTCTTG  5497

Qy   181  ACATTAACTTAACTATAAATGTTTATTTGATGAATTCAATGGCTAAGGTTG   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5496  ACATTAACTTAACTATAAATGTTTATTTGATGAATTCAATGGCTAAGGTTG  5437

Qy   241  GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGGCTAAGGTTGCAATTAA   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5436  GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGGCTAAGGTTGCAATTAA  5377
```

```
Qy  1381  CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4296  CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA  4237

Qy  1441  AGTTTGTTGGACTTCTGGCCATAATTGTGATCTCAAGACCAGATTCATTCTTAATAGCT  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4236  AGTTTGTTGGACTTCTGGCCATAATTGTGATCTCAAGACCAGATTCATTCTTAATAGCT  4177

Qy  1501  AAACAACAAACAAGAGATCCACAGGTTCAGCAGCTTATAAGAGTGAATTACTGATAC  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4176  AAACAACAAACAAGAGATCCACAGGTTCAGCAGCTTATAAGAGTGAATTACTGATAC  4117

Qy  1561  AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4116  AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC  4057

Qy  1621  ACAGAAGAATACGTACCATATGCCTGCATTTATGTGATATTCTAGCATTGTATTGTCCAA  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4056  ACAGAAGAATACGTACCATATGCCTGCATTTATGTGATATTCTAGCATTGTATTGTCCAA  3997

Qy  1681  CATAGTAGCCAGTAGCCTCACATGGCTATTCAAAATTTAAGTTGATTAAAATTAAGTAAGA  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3996  CATAGTAGCCAGTAGCCTCACATGGCTATTCAAAATTTAAGTTGATTAAAATTAAGTAAGA  3937

Qy  1741  ATAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAACTAGTGGCTACCACATCAGAC  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3936  ATAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAACTAGTGGCTACCACATCAGAC  3877

Qy  1801  GGTGCAAATATAGAATATTCCTTTATAACAGAAGTTCTATTGGAAAACAATGTTCTAG  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3876  GGTGCAAATATAGAATATTCCTTTATAACAGAAGTTCTATTGGAAAACAATGTTCTAG  3817

Qy  1861  AAAATATCACATATCTATAAAAACAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3816  AAAATATCACATATCTATAAAAACAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG  3757

Qy  1921  TGAGGGGAGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA  1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3756  TGAGGGGAGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA  3697

Qy  1981  ATCTTGATTGCGATGAAGGCTACTCCGGTGTCTAAATGTGTCACCTCCTCCAGACTGAACACT  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3696  ATCTTGATTGCGATGAAGGCTACTCCGGTGTCTAAATGTGTCACCTCCTCCAGACTGAACACT  3637

Qy  2041  TGGAATTGGCGAATTTCATTGTGAATTTATACCTCATAAAGTAACTCTAAGAGGTCA  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3636  TGGAATTGGCGAATTTCATTGTGAATTTATACCTCATAAAGTAACTCTAAGAGGTCA  3577

Qy  2101  AGTGTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3576  AGTGTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA  3517

Qy  2161  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTTGCTAGTCCTGAAAACAC  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3516  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTTGCTAGTCCTGAAAACAC  3457

Qy  2221  AGAGAGATGAATTAGAATAGCAAGCCTGCCCTCCAGCTGTCACAATCCAGTACAGGAGATG  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3456  AGAGAGATGAATTAGAATAGCAAGCCTGCCCTCCAGCTGTCACAATCCAGTACAGGAGATG  3397

Qy  2281  AGTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAAA  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3396  AGTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAAA  3337

Qy  2341  GTGCAGATATAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTTATGTGGAAAAA  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3336  GTGCAGATATAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTTATGTGGAAAAA  3277

Qy  2401  GAGGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCCAGAATGGAGTTT  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3276  GAGGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCCAGAATGGAGTTT  3217

Qy  2461  TAAAACATTCCTGTGGAGGGCAGAATAATGATCCAAGAGCAACCAGAAACCAGAAAAATA  2520
```

```
Qy   301  AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTTCAGGCCCAGCTTTACTTCCTAATCGAT   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5376  AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTTCAGGCCCAGCTTTACTTCCTAATCGAT  5317

Qy   361  GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5316  GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA  5257

Qy   421  CAGAACTTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5256  CAGAACTTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA  5197

Qy   481  CAGAAAAGTGTCATCAATTAAAATCACCAAACTCAGTCTTCAATTTGAGCAATAGT   540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5196  CAGAAAAGTGTCATCAATTAAAATCACCAAACTCAGTCTTCAATTTGAGCAATAGT  5137

Qy   541  TGGTGAATTTACTCCACCACCCTCCTCTGAAGGTTCTTCCTGCTCCTCCACTATA   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5136  TGGTGAATTTACTCCACCACCCTCCTCTGAAGGTTCTTCCTGCTCCTCCACTATA  5077

Qy   601  AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGTTCAGTTACCCTGACACAAAGGAAT   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5076  AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGTTCAGTTACCCTGACACAAAGGAAT  5017

Qy   661  TCAGTTCTCTGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT   720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5016  TCAGTTCTCTGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT  4957

Qy   721  CTGCTTAAGGCCTCTTGTGTGGCATCTGTTGTTTTCCACTGCGCCCTGTACCTATTGCTCT   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4956  CTGCTTAAGGCCTCTTGTGTGGCATCTGTTGTTTTCCACTGCGCCCTGTACCTATTGCTCT  4897

Qy   781  TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTTCTCGCTTCAGTCATGGTAGGG   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4896  TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTTCTCGCTTCAGTCATGGTAGGG  4837

Qy   841  CCATCAGTCACACATGATCAGGCCTCTCCTGCCAAACATGGCATCTTTCTTTGGGAATT   900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4836  CCATCAGTCACACATGATCAGGCCTCTCCTGCCAAACATGGCATCTTTCTTTGGGAATT  4777

Qy   901  TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTGAATCTGACTTACGCCTAC   960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4776  TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTGAATCTGACTTACGCCTAC  4717

Qy   961  AGTGGCTTTGCAAAGTGACTGTCCATTCCTATTTCTTAAGTCCCTGAATTTATAAATTTAT  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4716  AGTGGCTTTGCAAAGTGACTGTCCATTCCTATTTCTTAAGTCCCTGAATTTATAAATTTAT  4657

Qy  1021  CCTGGTTACAGCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCCTATAGTCTGT  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4656  CCTGGTTACAGCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCCTATAGTCTGT  4597

Qy  1081  GAATTTCATATTTCTTTCATACACATTTCATGTTTTGTTGTGTGTGTGTTTT  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4596  GAATTTCATATTTCTTTCATACACATTTCATGTTTTGTTGTGTGTGTGTTTT  4537

Qy  1141  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACA  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4536  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACA  4477

Qy  1201  TCCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4476  TCCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG  4417

Qy  1261  TCTGGGAGTAAGTAGTAGCCAGAGCTCGTACTAGCCTTAGCCAACTGCTTTACAC  1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4416  TCTGGGAGTAAGTAGTAGCCAGAGCTCGTACTAGCCTTAGCCAACTGCTTTACAC  4357

Qy  1321  AACATTTGCTCTCTCCTTCAGATAGCAGTCTTCGGAAGAAAGAGAGCTACTATTTTGC  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4356  AACATTTGCTCTCTCCTTCAGATAGCAGTCTTCGGAAGAAAGAGAGCTACTATTTTGC  4297
```

```
Db  2136 AGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAATAATAATAAATAAATAACTTTT 2077

Qy  3601 AAAAAACAAAAATTAATTAAATTTAAAAACACAACACTAGAGATGTTGCAAATTGA 3660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2076 AAAAAACAAAAATTAATTAAATTTAAAAACACAACACTAGAGATGTTGCAAATTGA 2017

Qy  3661 TTATTGGGAGTCTATATCCCTGGAGTTAATTTAAAATATTAGAGAGAGTTCTTCCTCA 3720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2016 TTATTTGGGAGTCTATATCCCTGGAGTTAATTTAAAATATTTAGAGAGAGTTCTTCCTCA 1957

Qy  3721 TTTCCTAGACGTCGAATTGTAAATATCAGAGCTAGAGGAACACTAGGGCTCGCCACT 3780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1956 TTTCCTAGACGTCGAATTGTAAATATCAGAGCTAGAAGGAACACTAGGGCTCGCCACT 1897

Qy  3781 CCAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAG 3840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1896 CCAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAG 1837

Qy  3841 AGTCTTAGGCCTCACCCCAGACTTAAAACTTGAGAAGCTCTGCACTAGAAATCTTCACTCCACCTTTC 3900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1836 AGTCTTAGGCCTCACCCCAGACCTCACTAGAAGCTCTGCACTAGAAATCTTCACTCCACCTTTC 1777

Qy  3901 GCCTCACGGCACATTAAAACTTGAGAAGCTCTGCACTAGAAATCTTCACTAGAAATTTCAGAA 3960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1776 GCCTCACGGCACATTAAAACTTGAGAAGCTCTGCACTAGAAATCTTCACTCCACCTTTC 1717

Qy  3961 ATTATAAAATGGAATCACTCGGCTGGTCACAGAAAATTGATTATTTTTAATTTCAGAA 4020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1716 ATTATAAATGGAATCACTCGGCTGGTCACAGAAAATTGATTATTTTTAATTTCAGAA 1657

Qy  4021 CCTTCTCATTTAGGTCATCTATATTTGCTAATAGCAGGGAAGAAAGCCAAACTCTTTAACT 4080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1656 CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGGAAGAAAGCCAAACTCTTTAACT 1597

Qy  4081 GCAATTAACAAAATCTATAATCTATAATTAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTTACATTTTG 4140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1596 GCAATTAACAAAATCTATAATCTATAATTAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG 1537

Qy  4141 TGGAGCAAGCTGTGTTGATTTGGCTGGGGCTCAGGCCGCTCAGGCCGCTCAGTGAATTCACAATT 4200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1536 TGGAGCAAGCTGTGTTGATTTGGCTGGGGCTCAGGCCGCTCAGGCCGCTCAGTGAATTTCACAATT 1477

Qy  4201 CACAGATGTTAGCCGCTCTGGGCTAAGTAAAGGAGAGAAGTGTCAAGTTTAAATAGCT 4260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1476 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAGAGAAGTGTCAAGTTTAAATAGCT 1417

Qy  4261 TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTATGAAACACATTTTGAGT 4320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1416 TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTTATGAAACACATTTTGAGT 1357

Qy  4321 TAGATTTACTTACACAGGGAAATGTCAAATTTCTTGAAAGGGGCTTCTGCTCCACAAC 4380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1356 TAGATTTACTTACACAGGGAAATGTCAAATTTCTTGAAAGGGGCTTCTGCTCCACAAC 1297

Qy  4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGTGAAGGGA 4440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1296 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGTGAAGGGA 1237

Qy  4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTTCCCCCTG 4500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1236 AGATGTGAACTCACCATGTTAGTGACCGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTTCCCCCTG 1177

Qy  4501 TTGGAGTCTATCCTAACTAACTGAGCTTCATTCATTCAATTTCAAATCATATTTCAATTTCCAAATCCACAAA 4560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1176 TTGGAGTCTATCCTAACTAACTGAGCTTCATTCATTCAATTTCAAATCATATTTCAATTTCCAAATCCACAAA 1117

Qy  4561 ACCAGGATAAGTTTACAGCCCATAATTCAGAAAAGGAATAAATATTTTGTGTGTAGACTT 4620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1116 ACCAGGATAAGTTTACAGCCCATAATTCAGAAAAGGAATAAATATTTTGTGTGTAGACTT 1057

Qy  4621 TCCTGATATTACACTGATTTGGGAATATATTGGAATATATGGAATATGTTTATGGTTTCCTTTGGAAGTAG 4680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1056 TCCTGATATTACACTGATTTGGGAATATATTGGAATATATGGAATATGTTTATGGTTTCCTTTGGAAGTAG 997
```

```
Db  3216 TAAAACATTCCTGGTTGGAGGGCAGAATATGATCCAAGGCAAGAGCAACCAGAAAAATA 3157
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2521 TGCAACCTAGAGGAAAGTGCATGAAGGGGAGCAGTTGTAAAATAATTTCATGATGATTAA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3097 TGCAACCTAGAGGAAAGTGCATGAAGGGGAGCAGTTGTAAAATAATTTCATGATGATGTAA 3156

Qy  2581 GTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCTCCTA 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3037 GTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCTCCTA 3096

Qy  2641 GGAGTCAAGAAGAACAAAGTGTCCCTTTCTTCCCTACGTTATGCTCAGTGGTCCAAGTCCAA 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2977 GGAGTCAAGAAGAACAAAGTGTCCCTTTCTTCCCTACGTTATGCTCAGTGGTCCAAGTCCAA 3036

Qy  2701 AACACCTTCCTTCCTTAAGTACTTTCTTCTTCCCCCATACAAATCTAAAGTCTTCACA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2917 AACACCTTCCTTCCTTAAGTACTTTCTTCCTTCCCCCATACAAATCTAAAGTCTTCACA 2976

Qy  2761 AACATCATTTAAAACGGCAGGTCATGGTCAGAAAGGCAATTGCTTTCCTAGACTTCTAT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2857 AACATCATTTAAAACAGGGCAGGTCATGGTCAGAAAGGCAATTGCTTTCCTAGACTTCTAT 2916

Qy  2821 GTACGTTATTATATTACAATTTCTGCCTAAAAGACTCTAAAAGTCTTGGAAAAGTTCCAC 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2797 GTACGTTATTATATTACAATTTCTGCCTAAAAGACTCTAAAAGTCTTGGAAAAGTTCCAC 2856

Qy  2881 CTTGCACATCAAAGATATAATCATGCATTGTATAGTAACCTTAGTCCCCTAAGAGAAT 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2737 CTTGCACATCAAAGATATAATCATGCATTGTATAGTAACCTTAGTCCCCTAAGAGAAT 2796

Qy  2941 AAGGATGAACTATAAATATAAGAAGTAATTATAATGGTAATTATTATAAATTGCCACTTATT 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2677 AAGGATGAACTATAATATAAGAAGTAATTATAATGGTAATTATTATAAATTGCCACTTATT 2736

Qy  3001 TTTCACTTGATCGTGTATGTCATGCTCGTTGCATGCTACTGTTCTGTTGAATTCTAGAGTTTG 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2617 TTTCACTTGATCGTGTATGTCATGCTCGTTGCATGCTACTGTTCTGTTGAATTCTAGAGTTTG 2677

Qy  3061 CCTCTTTTTCCTGGGTCAACTCTCGCCATTTATTTCCATAATGCAATAGAGAGCCAATCTT 3120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2557 CCTCTTTTTCCTGGGTCAACTCTCGCCATTTATTTCCATAATGCAATAGAGAGCCAATCTT 2616

Qy  3121 TTTCATAATACTTATTTAAAATTGTTGCCATTAATTTCTGTTCCTCTTAGCTTAGTA 3180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2497 TTTCATAATACTTATTTAAAATTGTTGCCATTAATTTCTGTTCCTCTTAGCTTAGTA 2556

Qy  3181 ACTTTAGGATTTTAAAATAACAACTAATTCATGACATACTACGTTAAATGTATTATTATT 3240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2437 ACTTTAGGATTTTAAAATAACAACTAATTCATGACATACTACGTTAAATGTATTATTATT 2496

Qy  3241 TAAATACGTTAGGCTATAAAACCTTTTAAAATTTTTAAAAATATGAGTGTGTGGGCT 3300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2377 TAAATACGTTAGGCTATAAAACCTTTTAAAATTTTTAAAAATAGAATGAGTGTGTGGGCT 2436

Qy  3301 CATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAG 3360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2317 CATGCCTGTAATCCCACACTTTGGGAAGGCCGGGTCGGGGAGGATAGCTTGAGTCCAGCAG 2376

Qy  3361 TTTGAGACCAGTCAGGGCCAACACAGCAAGAGACCCATATCTAAAAAACAAACAAAACAA 3420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2257 TTTGAGACCAGTCAGGGCCAACACAGCAAGAGACCCATATCTAAAAAACAAAACAAAACAA 2316

Qy  3421 AATTACCTGGTATGGTGTGCTCACCGTGTGTAGTCCAAGCTACACAGAAGCTGAGGCAGA 3480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2197 AATTACCTGGTATGGTGTGCTCACCGTGTGTAGTCCAAGCTACACAGAAGCTGAGGCAGA 2256

Qy  3481 AGGGATCACTTGAGCCCAGGAGGTTGAGGCGTTGAGGCTGCAGTGAGCTGCACTGCCATGAAGCGGCGCTACACTC 3540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2137 AGGGATCACTTGAGCCCAGGAGGTTGAGGCGTTGAGGCTGCAGTGAGCTGCACTGCCATGAACGCGGCGCTACACTC 2196

Qy  3541 AGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAATAAATAACTTTT 3600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

DEFINITION　Gene encoding promoter region of tumor suppressor gene p51.
ACCESSION　BD093092
VERSION　BD093092.1　GI:22638680
KEYWORDS　WO 0100818-A/4,.;
SOURCE　Homo sapiens (human)
ORGANISM　Homo sapiens
　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1 (bases 1 to 5676)
AUTHORS　Sakai.T., Kagaya.S., Sato.T., Sukenaga,Y. and Fujii.H.
TITLE　Gene encoding promoter region of tumor suppressor gene p51
JOURNAL　Patent: WO 0100818-A 4 04-JAN-2001;
　　　　TOSHIYUKI SAKAI,NIPPON KAYAKU KK,SHIGEHIDE KAGAYA,TAKAMICHI SATO,
　　　　YOSHIKAZU SUKENAGA,HIDEJI FUJII
OS　Homo sapiens (human)
COMMENT
PN　WO 0100818-A/4
PD　04-JAN-2001
PF　28-JUN-2000 WO 2000JP004261
PR　29-JUN-1999 JP 99P　183195
PI　TOSHIYUKI SAKAI,SHIGEHIDE KAGAYA,TAKAMICHI SATO,YOSHIKAZU　PI
　　SUKENAGA,
PI　HIDEJI FUJII
PC　C12N15/10,C12N5/10,C12Q1/68,A61K48/00,A61K45/00,A61P43/00,　PC
　　A61P35/00
CC
FH　Key　　Location/Qualifiers.
FEATURES　　Location/Qualifiers.
source　1..5676
　　　　　/organism="Homo sapiens"
　　　　　/mol_type="genomic RNA"
　　　　　/db_xref="taxon:9606"

ORIGIN

Query Match　　　100.0%;　Score 5676;　DB 6;　Length 5676;
Best Local Similarity　100.0%;　Pred. No. 0;
Matches 5676;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　　1 CAGCTGTTCAGGGATGCTCTGAGAAAAGAAGCCACCACATTGCTTCTGGACACTGGGTGT 60
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5676 CAGCTGTTCAGGGATGCTCTGAGAAAAGAAGCCACCACATTGCTTCTGGACACTGGGTGT 5617

Qy　　61 GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5616 GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 5557

Qy　121 TGGCCCTCGTTCCCTGCATGCCCTCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5556 TGGCCCTCGTTCCCTGCATGCCCTCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 5497

Qy　181 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGAGATGGAG 240
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5496 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGAGATGGAG 5437

Qy　241 GTCAAATCAGAAGAAGCACACATGGCTAAGGTTGCAATGCAAGGTTGCAATGCAATGCAA 300
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5436 GTCAAATCAGAAGAAGCACACATGGCTAAGGTTGCAATGCAATGCAATGCAATGCAATAA 5377

Qy　301 AGTCATTCGAATACCATTCAGTTACTTAAGTTCAGGCCCAGCTTTACTCCTAATCGAT 360
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5376 AGTCATTCGAATACCATTCAGTTACTTAAGTTCAGGCCCAGCTTTACTCCTAATCGAT 5317

Qy　361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGAGTTAGCAGAAGT 420
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5316 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGAGTTAGCAGAAGT 5257

Qy　421 CAGAGAGATTGGCTTCACTTTACTAAGATCCTTACTAAAATAGAGATTGTGTTGAGCTGA 480
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5256 CAGAGAGATTGGCTTCACTTTACTAAGATCCTTACTAAAATAGAGATTGTGTTGAGCTGA 5197

Qy　481 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTACCAGTCTCTTCAATTTGAGCAATAGT 540
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　5196 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTACCAGTCTCTTCAATTTGAGCAATAGT 5137

Qy　4681 GTCAAGTCAAAGCAAAACCAAAAAACAGCAAAAAACTGTAAGACATAAAAGAATAGAGTGGAG 4740
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　996 GTCAAGTCAAAGCAAAACCAAAAAACAGCAAAAAACTGTAAGACATAAAAGAATAGAGTGGAG 937

Qy　4741 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 4800
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　936 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 877

Qy　4801 GTGCACTTCAGAATATTCTACAATATATATATATTTTAATATCTTTAAGAAAA 4860
　　　　　|||||||||||||||||||||||||||||||||||||||||||
Db　876 GTGCACTTCAGAATATTCTACAATATATATATATTTTAATATCTTTAAGAAAA 817

Qy　4861 TTACTATATTATATGTAAGTACATGTGCATGTGTTGAGGTAGGATATTTAACTCAATAA 4920
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||
Db　816 TTACTATATTATATGTAAGTACATGTGCATGTGTTGAGGTAGGATATTTAACTCAATAA 757

Qy　4921 AGGTTATTTTCTTTTATTCGGGTCCTAAGGGTGTGAAAGGGATATCT 4980
　　　　　||||||||||||||||||||||||||||||||||||||||||||
Db　756 AGGTTATTTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGTGTGAAAGGGATATCT 697

Qy　4981 CTTTCTTTAGCTGAGAGGAAGAGTGAGTTCTAAGTAAATATAATAATCAAGGAATTTCCCT 5040
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　696 CTTTCTTTAGCTGAGAGGAAGAGTGAGTTCTAAGTAAATATAATAATCAAGGAATTTCCCT 637

Qy　5041 GTCTTTGCTATTTGAGATTGTGACCACACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 5100
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　636 GTCTTTGCTATTTGAGATTGTGACCACACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 577

Qy　5101 CGGGGAGGAGGGGAAATAGATAGATGAAAAAAACAAAACAAAAACTTCCCTCCCTAAGCAGCTCT 5160
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||
Db　576 CGGGGAGGAGGGGAAATAGATAGATGAAAAAAACAAAACAAAAACTTCCCTCCCTAAGCAGCTCT 517

Qy　5161 ACAAAACATTTTAGCCCCAGAAAGTCACAGAAATAGTCACAAATCCTCAAATCAAACCAGTATCCAGA 5220
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　516 ACAAAACATTTTAGCCCCAGAAAGTCACAGAAATAGTCACAAATCCTCAAATCAAACCAGTATCCAGA 457

Qy　5221 TACAAGGAAGTGTTATGTAGCTGAGCAGGTGGACACTTCAGTTCAGTTACA 5280
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　456 TACAAGGAAGTGTTATGTAGCTGAGCAGGTGGACACTTCAGTTCAGTTACA 397

Qy　5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAATCACGACA 5340
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　396 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAATCACGACA 337

Qy　5341 GAGATCAGAGAGTTCAGAGATGCCTCCAGCTCAAATTGCCAACAACAGAGTGTGGCTACTA 5400
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　336 GAGATCAGAGAGTTCAGAGATGCCTCCAGCTCAAATTGCCAACAACAAGTGTGGCTACTA 277

Qy　5401 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA 5460
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　276 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA 217

Qy　5461 GCTCTGGGTAAGAATCGAGTGTTTATGTAGAGTTTAGTCAATTGAGTTCAATTGATGAATCT 5520
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　216 GCTCTGGGTAAGAATCGAGTGTTTATGTAGAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA 157

Qy　5521 TCAAGAAACGCTCCGCCTCTTTGCAAATATGTGAAGGAGGAGAGTGCCTAAACTTCTA 5580
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　156 TCAAGAAACGCTCCGCCTCTTTGCAAATATGTGAAGGAGGAGAGTGCCTAAACTTCTA 97

Qy　5581 TGTTCTGATAGCATTTGTATATTTATATAATGGCCTTTTAGCCTTCCCCGGCTTTATATCTATATATACAC 5640
　　　　　||||||||||||||||||||||||||||||||||||||||||||||
Db　96 TGTTCTGATAGCATTTGTATATTTATATAATGGCCTTTTAGCCTTCCCCGGCTTTATATCTATATATACAC 37

Qy　5641 AGGTATTTGTGTATATTTTATATAATTGTTCTCCCGT 5676
　　　　　||||||||||||||||
Db　36 AGGTATTTGTGTATATTTTATATAATTGTTCTCCCGT 1

RESULT 3
BD093092/c
LOCUS　BD093092　　5676 bp　　RNA　　linear　　PAT 27-AUG-2002

```
Qy    541  TGGTGAATTACTCACCACCTCCTCCTCCTGAAGTTCTTCCTGCTCCTCCACTATA  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5136  TGGTGAATTACTCACCACCTCCTCCTCCTGAAGTTCTTCCTGCTCCTCCACTATA 5077

Qy    601  AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAGGAAT  660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5076  AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAGGAAT 5017

Qy    661  TCAGTTCTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCAGGCTGAATCGAAGATCAT  720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5016  TCAGTTCTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCAGGCTGAATCGAAGATCAT 4957

Qy    721  CTGCTTAAGCCTCTGTGTGGCATCTGTTGTTCCACTGCCCTGTACCTATTGCTCT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4956  CTGCTTAAGCCTCTGTGTGGCATCTGTTGTTCCACTGCCCTGTACCTATTGCTCT 4897

Qy    781  TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTGCTTGCTGAAAACTCTGC  840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4896  TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTGCTTGCTGAAAACTCTGC 4837

Qy    841  CCATCAGTCCACATGATCAGGCCCTCCTCGGCCAAACATGGCATCTTCTTTGGGAATT  900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4836  CCATCAGTCCACATGATCAGGCCCTCCTCGGCCAAACATGGCATCTTCTTTGGGAATT 4777

Qy    901  TGAATCTTAAGCTGAATAGCTCAAAAAAGCTGTGAATCGTGTCATTCCTATTTATTTAT 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4776  TGAATCTTAAGCTGAATAGCTCAAAAAAGCTGTGAATCGTGTCATTCCTATTTATTTAT 4717

Qy    961  AGTGGCTTTGCAAAGTGACTGTCCATTCCATTCCAATTCCTAAGTCCCTGAATTTAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4716  AGTGGCTTTGCAAAGTGACTGTCCATTCCATTCCAATTCCTAAGTCCCTGAATTTAT 4657

Qy   1021  CCTGGTTACAGCGCCCTTTCTGAGATGTGTGTTTTTTTGTTTGTTTGTGTGTTTTTT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4656  CCTGGTTACAGCGCCCTTTCTGAGATGTGTGTTTTTTTGTTTGTTTGTGTGTTTTTT 4597

Qy   1081  GAATTTCATATTTCTTCTTGTGTTTGTTGTTTATACCCAAGGAATCCTGATTGATACA 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4596  GAATTTCATATTTCTTCTTGTGTTTGTTGTTTATACCCAAGGAATCCTGATTGATACA 4537

Qy   1141  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTTGTTTATACCCAAGGAATCCTGATTGATACA 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4536  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTTGTTTATACCCAAGGAATCCTGATTGATACA 4477

Qy   1201  TCCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGTAGGTCGTACCTGCTTACAC 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4476  TCCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGTAGGTCGTACCTGCTTACAC 4417

Qy   1261  TCTGGGAGTAAGTTAGTACCAGAGTCGTACTAACCCCAGGTTAGCCAACTGCTTATTTGC 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4416  TCTGGGAGTAAGTTAGTACCAGAGTCGTACTAACCCCAGGTTAGCCAACTGCTTATTTGC 4357

Qy   1321  AACATTTGCTCTCCTTCCTCAGATGCTTATAGCAGTCTTGGAGAAAGAAGCTACTATTTTGC 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4356  AACATTTGCTCTCCTTCCTCAGATGCTTATAGCAGTCTTGGAGAAAGAAGCTACTATTTTGC 4297

Qy   1381  CAAAGACCTCAGGAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGAGTGAACTCTTAAAA 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4296  CAAAGACCTCAGGAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGAGTGAACTCTTAAAA 4237

Qy   1441  AGTTTGTTGGACTTCTGGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4236  AGTTTGTTGGACTTCTGGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT 4177

Qy   1501  AAACAAACAAACAAGAGATCCACAGGTTCAGCAGTTCACCAGTTGTATCATAAGAGTGAATTACTGATAC 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4176  AAACAAACAAACAAGAGATCCACAGGTTCAGCAGTTCACCAGTTGTATCATAAGAGTGAATTACTGATAC 4117

Qy   1561  AGTTGACAACATGAATATCTCAGAAAACCATGGCATCAATGAGCAATGAGAAAAAATCCAGAC 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4116  AGTTGACAACATGAATATCTCAGAAAACCATGGCATCAATGAGCAATGAGAAAAAATCCAGAC 4057

Qy   1621  ACAGAAGAATACGTACCATATGCCTGCATTTATGTGATATTCTAGCATTGTATTGTCCAA 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    600  ACAGAAGAATACGTACCATATGCCTCACATGTTCATGCATTTATGTGATATTCTAGCATTGTATTGTCCAA 3997
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4056  ACAGAAGAATACGTACCATATGCCTCACATGTTCATGCATTTATGTGATATTCTAGCATTGTATTGTCCAA 3997

Qy   1681  CATAGTAGCCAGTAGCCTCAGTAGCTTCAGTGTAACTAGTGGCTACCACCATCAGAC 1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3996  CATAGTAGCCAGTAGCCTCAGTAGCTTCAGTGTAACTAGTGGCTACCACCATCAGAC 3937

Qy   1741  ATAAAAATTTAGCTCTTCAGTAGCTTAGTCTTCAGTAGCGTTAGCCACCATGTAACTAGTGGCTACCACCATCAGAC 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3936  ATAAAAATTTAGCTCTTCAGTAGCTTAGTCTTCAGTAGCGTTAGCCACCATGTAACTAGTGGCTACCACCATCAGAC 3877

Qy   1801  GGTGCAAATATAGAATATTTCCTTTATAACAGAGAAAGTTCATTGAAAACAATGTTCTAG 1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3876  GGTGCAAATATAGAATATTTCCTTTATAACAGAGAAAGTTCATTGAAAACAATGTTCTAG 3817

Qy   1861  AAAATATACACATAATCTATAAAACAAAAGCAAGTCAGTGATTGTCTCAAGGCCAGGGG 1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3816  AAAATATACACATAATCTATAAAACAAAAGCAAGTCAGTGATTGTCTCAAGGCCAGGGG 3757

Qy   1921  TGAGGGGAGATCGATTGCAAAGTGTATGAGGAAAGTTTTGGGGTAATAGGGTTGTTGGA 1980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3756  TGAGGGGAGATCGATTGCAAAGTGTATGAGGAAAGTTTTGGGGTAATAGGGTTGTTGGA 3697

Qy   1981  ATCTTGATTGCGATGAAGAGGCTACTCCGGTGTCTAATGTGTCACCTCCTCGAGACTGCAGACACT 2040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3696  ATCTTGATTGCGATGAAGAGGCTACTCCGGTGTCTAATGTGTCACCTCCTCGAGACTGCAGACACT 3637

Qy   2041  TGGAATTGGGCGAATTTCATTGTATGTAAAATATACCTCATAAAGTAACTCTAAGAGTCA 2100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3636  TGGAATTGGGCGAATTTCATTGTATGTAAAATATACCTCATAAAGTAACTCTAAGAGTCA 3577

Qy   2101  AGTGTTTTGTGTGGAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA 2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3576  AGTGTTTTGTGTGGAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA 3517

Qy   2161  AATACATAAACATGTATTCATCCATTAGGTGCAATATTTTGCAATATTTTGCGAAAACAC 2220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3516  AATACATAAACATGTATTCATCCATTAGGTGCAATATTTTGCAATATTTTGCGAAAACAC 3457

Qy   2221  AGAGATGAATTAGAATAGCAAGCCTGCCCCGCCAGGAAGAAGTGGGAAGGTTTCTTTATGCGACGAGATG 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3456  AGAGATGAATTAGAATAGCAAGCCTGCCCCGCCAGGAAGAAGTGGGAAGGTTTCTTTATGCGACGAGATG 3397

Qy   2281  AGTCTATTCAAAAGTAATAGCTAGACTCCAGGAAGAAGTGTATAGGTGACCTTACACAAAAA 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3396  AGTCTATTCAAAAGTAATAGCTAGACTCCAGGAAGAAGTGTATAGGTGACCTTACACAAAAA 3337

Qy   2341  GTGCAGATATAATTATGTAGGACACACCTGAGTTTGGCAGAAGTCATGTCTTATGTGGAAAAA 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3336  GTGCAGATATAATTATGTAGGACACACCTGAGTTTGGCAGAAGTCATGTCTTATGTGGAAAAA 3277

Qy   2401  GAGGGAGAATTTTGTCTTTGAAGGATGAGCAAGATGCAGATGGAGTTT 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3276  GAGGGAGAATTTTGTCTTTGAAGGATGAGCAAGATGCAGATGGAGTTT 3217

Qy   2461  TAAAACATTCCTGGTGGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATA 2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3216  TAAAACATTCCTGGTGGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATA 3157

Qy   2521  TGCAACCTAGAGGAGAAGTGCATGAGAGGGGAGCAGTTGTAAAATAATTTCATGAATGTAA 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3156  TGCAACCTAGAGGAGAAGTGCATGAGAGGGGAGCAGTTGTAAAATAATTTCATGAATGTAA 3097

Qy   2581  GTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGTGTCATGGCGTTCTTGGCCTCCTA 2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3096  GTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGTGTCATGGCGTTCTTGGCCTCCTA 3037

Qy   2641  GGAGTCAAGAAGAACAAAGTGTCCCTTTCTCCCCAGTGCTCAGTGGTCCAAGTCCAA 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3036  GGAGTCAAGAAGAACAAAGTGTCCCTTTCTCCCCAGTGCTCAGTGGTCCAAGTCCAA 2977

Qy   2701  AACACCTTTCCTTCCTTCCTTAAGTACTTTCTTCTCCCCATACAAAATCTAAAGTCTTCACA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db 2976 AACACCTTTCCTTCCTTAAGTACTTCTTCTCCCCTCCATACAAATCTAAAGTCTTCACA 2917
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2761 AACATCATTTAAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTCCTAGACTTCTAT 2820

Db 2916 AACATCATTTAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTCCTAGACTTCTAT 2857
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2821 GTACGTTATTATATTACAATTTCTGCCTAAAGACTCTAAAGTCTTGGAAAAGTTCCAC 2880

Db 2856 GTACGTTATTATATTACAATTTCTGCCTAAAGACTCTAAAGTCTTGGAAAAGTTCCAC 2797
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 2881 CTTGCACATCAAAGATATAATTCATGCATTGTATAGTAGTAACCTTAGTCCCCTAAGAGAAT 2940

Db 2796 CTTGCACATCAAAGATATAATTCATGCATTGTATAGTAGTAACCTTAGTCCCCTAAGAGAAT 2737
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2941 AAGGATGAACTATAAAATATAGAAGTAATATGGTAATTATAATGATTGCCACTTATT 3000

Db 2736 AAGGATGAACTATAAAATATAGAAGTAATATGGTAATTATAATGATTGCCACTTATT 2677
         ||||||||||||||||||||||||||||||||||||||||||
Qy 3001 TTTCACTTGATCGTGTATGGTGCATGCTACTGGTGTTCTGTTCTGTTGAATTCAGAGAGTTTG 3060

Db 2676 TTTCACTTGATCGTGTATGGTGCATGCTACTGGTGTTCTGTTCTGTTGAATTCAGAGAGTTTG 2617
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 3061 CCTCTTTTTCCTGGGTCAACTCTCGCCCATTTATTTCCATAATGCAATAGAGCCAATCTT 3120

Db 2616 CCTCTTTTTCCTGGGTCAACTCTCGCCCATTTATTTCCATAATGCAATAGAGCCAATCTT 2557
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3121 TTTCATAATTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA 3180

Db 2556 TTTCATAATTACTTATTTAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA 2497
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3181 ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATT 3240

Db 2496 ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATT 2437
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3241 TAAATACGTTAGGCTATAAAACCTTTTAAATTTTTAAAAAATAGAGATGAGTGTGGTGGCT 3300

Db 2436 TAAATACGTTAGGCTATAAAACCTTTTAAATTTTTAAAAAATAGAGATGAGTGTGGTGGCT 2377
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3301 CATGGCTGTAATCCCAACACTTGGGAAGCCGGTCGGGAGGAGTAGCTTGAGTCCAGCAG 3360

Db 2376 CATGGCTGTAATCCCAACACTTGGGAAGCCGGTCGGGAGGAGTAGCTTGAGTCCAGCAG 2317
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3361 TTTGAGACCAGTCAGGGCAACACAGCAAGAGCCCCATCTAAAAAACAAAACAAAACAA 3420

Db 2316 TTTGAGACCAGTCAGGGCAACACAGCAAGAGCCCCATCTAAAAAACAAAACAAAACAA 2257
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3421 AATTACCTGGGTATGGTTGTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGA 3480

Db 2256 AATTACCTGGGTATGGTTGTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGA 2197
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3481 AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTCGATCCATGAACGCGCTCTACCACTC 3540

Db 2196 AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTCGATCCATGAACGCGCTCTACCACTC 2137
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3541 AGTCTCGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAATAAAATAACTTTT 3600

Db 2136 AGTCTCGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAATAAAATAACTTTT 2077
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 3601 AAAAAACAAAAATTAATTAATTTAAAAACCAACACAACAACAACACTAGAGATATTTAGAAT 3660

Db 2076 AAAAAACAAAAATTAATTAATTTAAAAACCAACACAACAACAACACTAGAGATATTTAGAAT 2017
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3661 TTATTTGGGAGTCTATATCCCTGGAGTTAAATATCAGAGTCAAGGAGTTCTTCCTCA 3720

Db 2016 TTATTTGGGAGTCTATATCCCTGGAGTTAAATATCAGAGTCAAGGAGTTCTTCCTCA 1957
         ||||||||||||||||||||||||||||||||||||||||||
Qy 3721 TTTCCTAGAGACGTCGAATTGTAAATATCAGAGTGAAGGAACACTAGGGCTCGCCACT 3780

Db 1956 TTTCCTAGAGACGTCGAATTGTAAATATCAGAGTGAAGGAACACTAGGGCTCGCCACT 1897
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 3781 CCAAAGTGTGGTCCAAGGACCATCAAGTCAAGTAACCTGGGAACGTGTTAGAAATGCAG 3840

Db 1896 CCAAAGTGTGGTCCAAGGACCATCAAGTCAAGTAACCTGGGAACGTGTTAGAAATGCAG 1837

Qy 3841 AGTCTTAGGCCTCACCCCAGACCTACTGAACCAGAATCTGCATTAACAGATTCTAGGT 3900

Db 1836 AGTCTTAGGCCTCACCCCAGACCTACTGAACCAGAATCTGCATTAACAGATTCTAGGT 1777
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 3901 GCCTCACGGCACATTAAAACTTGAGAAGCTCTGACTACTAGAAATCTTCACTCCACCTTTC 3960

Db 1776 GCCTCACGGCACATTAAAACTTGAGAAGCTCTGACTACTAGAAATCTTCACTCCACCTTTC 1717
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 3961 ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATTGATTATTTTTAATTTCAGAA 4020

Db 1716 ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATTGATTATTTTTAATTTCAGAA 1657
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4021 CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGTGGAGAAAGCCAAACTCTTTAACT 4080

Db 1656 CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGTGGAGAAAGCCAAACTCTTTAACT 1597
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4081 GCAATTAACAAATCTATAATTAATTAATTAGTTAAGCAATCTTCCCTTAAGTTTTACATTTTG 4140

Db 1596 GCAATTAACAAATCTATAATTAATTAATTAGTTAAGCAATCTTCCCTTAAGTTTTACATTTTG 1537
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4141 TGGAGCAAGCTGTTTGGCTGGGCTCAGGCGCCGGCCTGTTTGTGAATTTCACAATT 4200

Db 1536 TGGAGCAAGCTGTTTGGCTGGGCTCAGGCGCCGGCCTGTTTGTGAATTTCACAATT 1477
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 4201 CACAGATGTTAGCCGCTCGGGCTCAAGTAAAGGAAGGAGAGAATGTCAAGTTTAAATAGCT 4260

Db 1476 CACAGATGTTAGCCGCTCGGGCTCAAGTAAAGGAAGGAGAGAATGTCAAGTTTAAATAGCT 1417
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4261 TCTCCCTTCCATCCTGCTGAAGCAACAACAAATAAATATTTTTATGGAAACATTTTGAGT 4320

Db 1416 TCTCCCTTCCATCCTGCTGAAGCAACAACAAATAAATATTTTTATGGAAACATTTTGAGT 1357
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4321 TAGAATTTACTTACAGGGAAATGTCAAATTGTCTCATTACATGAGCTTACATCATATTT 4380

Db 1356 TAGAATTTACTTACAGGGAAATGTCAAATTGTCTCATTACATGAGCTTACATCATATTT 1297
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 4381 TTTGACATCTACATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGAGGGGTGAAGGGA 4440

Db 1296 TTTGACATCTACATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGAGGGGTGAAGGGA 1237
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGTGACATACACAGAGTGGTTTTTTCCCCTG 4500

Db 1236 AGATGTGAACTCACCATGTTAGTGACCGTTAGTGACATACACAGAGTGGTTTTTTCCCCTG 1177
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4501 TTGGAGTCTATCCTAAACTGAGCTTCTGAATCATATTTCATTCCAAATCCACAAA 4560

Db 1176 TTGGAGTCTATCCTAAACTGAGCTTCTGAATCATATTTCATTCCAAATCCACAAA 1117
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 4561 ACCAGGATAAGTTACAGCCCATAATTCAGAAAGGAAATAAATATTTTGTGTAGACTT 4620

Db 1116 ACCAGGATAAGTTACAGCCCATAATTCAGAAAGGAAATAAATATTTTGTGTAGACTT 1057
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 4621 TCCTGATATTACACTGATTTGGGAATATATGGAATAACAATTTATGGTTTCCTTTCGAAGTAG 4680

Db 1056 TCCTGATATTACACTGATTTGGGAATATATGGAATAACAATTTATGGTTTCCTTTCGAAGTAG 997
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4681 GTCAAGTCAAAGCAAAAGCAAAACCAAAAACAGCAAAACTGTAAGACATAAAGAATAGAGTGGAG 4740

Db 996 GTCAAGTCAAAGCAAAAGCAAAACCAAAAACAGCAAAACTGTAAGACATAAAGAATAGAGTGGAG 937
         ||||||||||||||||||||||||||||||||||||||||||||||
Qy 4741 CCGACTGAGAGATTAAAAATAAAACTAGAAATATTTGAAATAATTT 4800

Db 936 CCGACTGAGAGATTAAAAATAAAACTAGAAATATTTGAAATAATTT 877
         ||||||||||||||||||||||||||||||||||||||||||
Qy 4801 GTGCACTTCAGAATATTCTACAATAATAATATATTCCAATTTAATATCTTTAAGAAAA 4860

Db 876 GTGCACTTCAGAATATTCTACAATAATAATATATTCCAATTTAATATCTTTAAGAAAA 817
         ||||||||||||||||||||||||||||||||||||||||||||
Qy 4861 TTACTATTATTATGTAAGTACATGTGCATGTGTTGAGTAGGATATTTAACTCAATTAA 4920

Db 816 TTACTATTATTATGTAAGTACATGTGCATGTGTTGAGTAGGATATTTAACTCAATTAA 757

```
PF   28-JUN-2000 WO 2000JP004261
PR   29-JUN-1999 JP 99P  183195
PI   TOSHIYUKI SAKAI,SHIGEHIDE KAGAYA,TAKAMICHI SATO,YOSHIKAZU        PI
     SUKENAGA,
PI   HIDEJI FUJII
PC   C12N15/10,C12N5/10,C12Q1/68,A61K48/00,A61K45/00,A61P43/00,       PC
     A61P35/00
CC
FH   Key             Location/Qualifiers
FT   promoter        (1)..(5676)
FT   5'UTR           (5677)..(5960)
FT   intron          (5677)..(5960).
FEATURES             Location/Qualifiers
     source          1..5960
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             100.0%;    Score 5676;   DB 6;   Length 5960;
Best Local Similarity   100.0%;    Pred. No. 0;
Matches 5676;           Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy      1 CAGCTGTTCAGGGATGCTCTGGAAAAGAAGCCCACCCACATTGCTTCTGGACACTGGGTGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CAGCTGTTCAGGGATGCTCTGGAAAAGAAGCCCACCCACATTGCTTCTGGACACTGGGTGT 60

Qy     61 GACTTTGGAGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACTTTGGAGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120

Qy    121 TGGCCTCTGTTCCCTGCATGCCCTCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGCCCCTGTTCCCTGCATGCCCTCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180

Qy    181 ACATTAACTAACTAACTATAAATGTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACATTAACTAACTAACTATAAATGTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 240

Qy    241 GTCAAATCAGAAGAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTCATTGAATTAA 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCAAATCAGAAGAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTCATTGAATTAA 300

Qy    301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGCTTTACTCCTAATCGAT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGCTTTACTCCTAATCGAT 360

Qy    361 GTCAGACTGTAGCAAATATTAGGTCAAAGTTGAAGAGTTAGCAGGATCCTCCATGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTCAGACTGTAGCAAATATTAGGTCAAAGTTGAAGAGTTAGCAGGATCCTCCATGA 420

Qy    421 CAGAACTTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGTGTGGTTGTGGCTATGTA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CAGAACTTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGTGTGGTTGTGGCTATGTA 480

Qy    481 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTCTTCAATTTGAGCAATAGT 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTCTTCAATTTGAGCAATAGT 540

Qy    541 TGGTGAATTTACTCCACCACCTCCTCCTTCCTGCTCCTCCTCCTGAAGGTTCTTTCCTGCTCCTCCTCCTGAAGGTTCTTTCCTGCTCCTCCTCCTCCTCCTGAAGGTTCTTTCCTGCTCCTCCTCCTCCTCCTCCTGAAGGTTCTTTCCTGCTCCTCCTCACTATA 600
          |||||||||||||||||||||||||||||||
Db    541 TGGTGAATTTACTCCACCACCTCCTCCTTCCTGCTCCTCCTCCTGAAGGTTCTTTCCTGCTCCTCCTCACTATA 600

Qy    601 AATGCAGGATGACCTGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAAGGAAT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATGCAGGATGACCTGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAAGGAAT 660

Qy    661 TCAGTTTCTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCTACGGAACATGCAAGATCAT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCAGTTTCTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCTACGGAACATGCAAGATCAT 720

Qy    721 CTGCTTTAAGCCTCTTGTGGTGGCATCTGTGTTTCCACTGCCCGTGTACCTATTGCTCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   4921 AGGTATTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCT 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    756 AGGTATTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCT 697

Qy   4981 CTTTCTTAGCTGAGAGGAAGAGTGAGTTCTAAGTTAAATATAATCAAGGAATTCCCT 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    696 CTTTCTTAGCTGAGAGGAAGAGTGAGTTCTAAGTTAAATATAATCAAGGAATTCCCT 637

Qy   5041 GTCTTTGCTATTGGAGATTGTGACCACAACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 5100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 GTCTTTGCTATTGGAGATTGTGACCACAACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 577

Qy   5101 CGGGAGGGAGGGGAAATAGATAGAAAACAAAACAAAACTTCCCTAAGCAGCTCT 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 CGGGAGGGAGGGGAAATAGATAGAAAACAAAACAAAACTTCCCTAAGCAGCTCT 517

Qy   5161 ACAAAACATTTAGCCCCAGAAAATCCTCAAATCCTCAAATCAAACCAGTATCCAGA 5220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 ACAAAACATTTAGCCCCAGAAAATCCTCAAATCCTCAAATCAAACCAGTATCCAGA 457

Qy   5221 TACAAGGAAGTGTTATGTAGCTGGAGCAGGGTGGACACTCACTCAGTTCAGTTACA 5280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    456 TACAAGGAAGTGTTATGTAGCTGGAGCAGGGTGGACACTCACTCAGTTCAGTTACA 397

Qy   5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCAGCATCCAATCACGACA 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    396 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCAGCATCCAATCACGACA 337

Qy   5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTA 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    336 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTA 277

Qy   5401 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGGATGCCCA 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGGATGCCCA 217

Qy   5461 GCTGGTAAGAATCGAGTGTTTAGTGTTGTTAGTCAATTGAATCTCATTGGCTAAAA 5520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    216 GCTGGTAAGAATCGAGTGTTTAGTGTTGTTAGTCAATTGAATCTCATTGGCTAAAA 157

Qy   5521 TCAAGAAAACCGTCCCGCCTCTTGCAAATATGTGAAGAGAGAGTGCCTAAACTTCTA 5580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 TCAAGAAAACCGTCCCGCCTCTTGCAAATATGTGAAGAGAGAGTGCCTAAACTTCTA 97

Qy   5581 TGTCTGATAGCATTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATACAC 5640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     96 TGTCTGATAGCATTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATACAC 37

Qy   5641 AGGTATTTGTGTATATTTATATAATTGTTCTCCGT 5676
          |||||||||||||||||||||||||||||||||||
Db     36 AGGTATTTGTGTATATTTATATAATTGTTCTCCGT 1
```

```
RESULT 4
BD093090
LOCUS       BD093090             5960 bp    DNA     linear    PAT 27-AUG-2002
DEFINITION  Gene encoding promoter region of tumor suppressor gene p51.
ACCESSION   BD093090
VERSION     BD093090.1  GI:22638678
KEYWORDS    WO 0100818-A/2.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5960)
AUTHORS     Sakai,T., Kagaya,S., Sato,T., Sukenaga,Y. and Fujii.H.
TITLE       Gene encoding promoter region of tumor suppressor gene p51
JOURNAL     Patent: WO 0100818-A 2 04-JAN-2001;
            TOSHIYUKI SAKAI,NIPPON KAYAKU KK,SHIGEHIDE KAGAYA,TAKAMICHI SATO,
            YOSHIKAZU SUKENAGA,HIDEJI FUJII
COMMENT     OS   Homo sapiens (human)
            PN   WO 0100818-A/2
            PD   04-JAN-2001
```

```
Db   721  CTGCTTTAAGCCCTCTTGTGGTGGCATCGTGTTTCCACTGCCCGTACCTATTGCTCT       780

Qy   781  TTCCTTGGTTAACAGAACCTTTATTTCTTCTGAAAACTTCTCTGCCTCAGTCATGGTAGGG   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  TTCCTTGGTTAACAGAACCTTTATTTCTTCTGAAAACTTCTCTGCCTCAGTCATGGTAGGG   840

Qy   841  CCATCAGTCCACATGATCAGGCCCTCCTCCTGGCCAAACATGGCATCTTTCTTTTGGGAATT  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  CCATCAGTCCACATGATCAGGCCCTCCTCCTGGCCAAACATGGCATCTTTCTTTTGGGAATT  900

Qy   901  TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTAC    960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTAC    960

Qy   961  AGTGGCTTTGCAAAGTGACTGTGTCCATTCCATTCCTATTTCTTAAGTCCCTGAATTTATATAATTTAT  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AGTGGCTTTGCAAAGTGACTGTGTCCATTCCATTCCTATTTCTTAAGTCCCTGAATTTATATAATTTAT  1020

Qy  1021  CCTGGTTACAGCCCCTTTCTGAGATGTGTGTTTTTTTTCCAACTGTCTCTTATAGTCTGT         1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CCTGGTTACAGCCCCTTTCTGAGATGTGTGTTTTTTTTCCAACTGTCTCTTATAGTCTGT         1080

Qy  1081  GAATTTCATATTTCTTTTCATACATTTCATGTTTGTTTGTTTGTTTGTTTGTTTGTTTTTT        1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GAATTTCATATTTCTTTTCATACATTTCATGTTTGTTTGTTTGTTTGTTTGTTTGTTTTTT        1140

Qy  1141  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACA       1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  GGCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACA       1200

Qy  1201  TCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG        1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  TCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG        1260

Qy  1261  TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTACAACCCCAGGTTAGCCAACTGCTTTACAC    1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTACAACCCCAGGTTAGCCAACTGCTTTACAC    1320

Qy  1321  AACATTTGCTTCTCTCCCTCCAGAGTTATAGCAGTCTTGGAGAAGAAGCTACTATTTTGC      1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  AACATTTGCTTCTCTCCCTCCAGAGTTATAGCAGTCTTGGAGAAGAAGCTACTATTTTGC      1380

Qy  1381  CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA     1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA     1440

Qy  1441  AGTTTGTTGACTTCTCGGCCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT    1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  AGTTTGTTGACTTCTCGGCCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT    1500

Qy  1501  AAACAAACAACAAGAGATATCTCAGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAATCCAGAC  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  AAACAAACAACAAGAGATATCTCAGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAATCCAGAC  1560

Qy  1561  AGTTGACAACATGAATATCTCAGAATATCTCAGAAACCATGAGCAATGAGAAAAAATCCAGAC  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  AGTTGACAACATGAATATCTCAGAATATCTCAGAAACCATGAGCAATGAGAAAAAATCCAGAC  1620

Qy  1621  ACAGAAGAATACGTACCTACCATATGCCTGCATTTATGTGATATTCTAGACATTGTATTGTCCAA  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  ACAGAAGAATACGTACCTACCATATGCCTGCATTTATGTGATATTCTAGACATTGTATTGTCCAA  1680

Qy  1681  CATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTAAGTTAAGTAAGATTAAGTAAGA  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  CATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTAAGTTAAGTAAGATTAAGTAAGA  1740

Qy  1741  ATAAAAATTTAGCTCTTCAGTAGCGTTAGCGTTAACTAGTGGCTACCACCATCAGAC         1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  ATAAAAATTTAGCTCTTCAGTAGCGTTAGCGTTAACTAGTGGCTACCACCATCAGAC         1800

Qy  1801  GGTGCAAATATAGAATATTTCCTTTATAACAGAAAAGTTCTATTGGAAAAACAATGTTCTAG    1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  GGTGCAAATATAGAATATTTCCTTTATAACAGAAAAGTTCTATTGGAAAAACAATGTTCTAG    1860
```

```
Qy  1861  AAAATATACACATAATCATATAAAAACAAAAAGCAGTCAGTCAGTGATTGTCTAAGGCCAGGGG  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  AAAATATACACATAATCATATAAAAACAAAAAGCAGTCAGTCAGTGATTGTCTAAGGCCAGGGG  1920

Qy  1921  TGAGGGGAGATCGATTGCAAAGTGGTATGGAGAAAGTTTGGGTAATAGGGTTGTTGGA       1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  TGAGGGGAGATCGATTGCAAAGTGGTATGGAGAAAGTTTGGGTAATAGGGTTGTTGGA       1980

Qy  1981  ATCTTGATTGCGATGAAGGCTACTCCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT    2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  ATCTTGATTGCGATGAAGGCTACTCCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT    2040

Qy  2041  TGGAATTGGCGAATTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGGAGGTCA     2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  TGGAATTGGCGAATTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGGAGGTCA     2100

Qy  2101  AGTGTTTGTTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA      2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  AGTGTTTGTTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA      2160

Qy  2161  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTCCTGAAAACAC     2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTCCTGAAAACAC     2220

Qy  2221  AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGAGGAGATG   2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGAGGAGATG   2280

Qy  2281  AGTCTATTCAAAAAATAGCTAGACTCCAGGAGAAAAGTTATAGGTGACCTTACACAAAAAA    2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  AGTCTATTCAAAAAATAGCTAGACTCCAGGAGAAAAGTTATAGGTGACCTTACACAAAAAA    2340

Qy  2341  GTGCAGATATAATTATGTGAGACAGTAGAAGTGGGAAGGTTTCTTTTATGTGGAAAAA       2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  GTGCAGATATAATTATGTGAGACAGTAGAAGTGGGAAGGTTTCTTTTATGTGGAAAAA       2400

Qy  2401  GAGGGAGAATTTTGGTCTTTGAAGGATGAGACGCAGAATGCGCAGATGGAGTTT          2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GAGGGAGAATTTTGGTCTTTGAAGGATGAGACGCAGAATGCGCAGATGGAGTTT          2460

Qy  2461  TAAAACATTCCTGGTGTGGAGGGCAGAATATGATCCAAGGCACACAGAGCAACCAGAAAAATA   2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  TAAAACATTCCTGGTGTGGAGGGCAGAATATGATCCAAGGCACACAGAGCAACCAGAAAAATA   2520

Qy  2521  TGCAACCTAGAGGAGAAAGTGCATGAAGGGGAGCAGTTGTAAAATAATTTCATGAATGTAA    2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TGCAACCTAGAGGAGAAAGTGCATGAAGGGGAGCAGTTGTAAAATAATTTCATGAATGTAA    2580

Qy  2581  GTGAGAAGAATTTGTATCATAGACACCTGAGTTGGCAGTGCATGTTCTTGGCTCCTA        2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  GTGAGAAGAATTTGTATCATAGACACCTGAGTTGGCAGTGCATGTTCTTGGCTCCTA        2640

Qy  2641  GGAGTCAAGAAGAACAAAGTGCCCTTCTCCCCCTTCTCTCTGGTCAGAAAGGCAAGTCCAA    2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  GGAGTCAAGAAGAACAAAGTGCCCTTCTCCCCCTTCTCTCTGGTCAGAAAGGCAAGTCCAA    2700

Qy  2701  AACACCTTCCTTCCTTAAGTACTTTCTCCATACAAATCTAAAGTCTTCACA             2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  AACACCTTCCTTCCTTAAGTACTTTCTCCATACAAATCTAAAGTCTTCACA             2760

Qy  2761  AACATCATTTAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTTCCTAGACTTCTAT      2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  AACATCATTTAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTTCCTAGACTTCTAT      2820

Qy  2821  GTACGTTATTATATTACCAATTCTCGCCTAAAAGACTCTAAAGTTTCCAC               2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  GTACGTTATTATATTACCAATTCTCGCCTAAAAGACTCTAAAGTTTCCAC               2880

Qy  2881  CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAACCTTAGTCCCCTAAGAGAAT    2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAACCTTAGTCCCCTAAGAGAAT    2940
```

```
Qy   2941  AAGGATGAACTATAAATATAAGAAGTAATTATGGTAATTATAATATGATTGCCACTTATT  3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  AAGGATGAACTATAAATATAAGAAGTAATTATGGTAATTATAATATGATTGCCACTTATT  3000

Qy   3001  TTTCACTTGATCGTGTATGGTTCATGCTGGTCTGTTCTGGTTCTTGAATTCTAGAGAGTTTG  3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001  TTTCACTTGATCGTGTATGGTTCATGCTGGTCTGTTCTGGTTCTTGAATTCTAGAGAGTTTG  3060

Qy   3061  CCTCTTTTCCTGGGTCAACTCTCGCCATTTATTCATAAATGCAATAGGAGCCAATCTT  3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  CCTCTTTTTCCTGGGTCAACTCTCGCCATTTATTCATAAATGCAATAGGAGCCAATCTT  3120

Qy   3121  TTTCATAATTACTTATTTATTAAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA  3180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121  TTTCATAATTACTTATTTATTAAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA  3180

Qy   3181  ACTTTAGGATTTTTAAATAACAACTATTGAAAATCATGACATACGTTAAATGATATTATT  3240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  ACTTTAGGATTTTTAAATAACAACTATTGAAAATCATGACATACGTTAAATGATATTATT  3240

Qy   3241  TAAATACGTTAGGCTATAAAACCTTTTAAAATTTTTAAAAAAATAGATGAGTGTGGTGGCT  3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241  TAAATACGTTAGGCTATAAAACCTTTTAAAATTTTTAAAAAAATAGATGAGTGTGGTGGCT  3300

Qy   3301  CATGCCTGTAATCCCAACACTTTGGGAAGCCAACAGCAAGACCCCATATCTAAAAAACAAAACAA  3360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  CATGCCTGTAATCCCAACACTTTGGGAAGCCAACAGCAAGACCCCATATCTAAAAAACAAAACAA  3360

Qy   3361  TTTGAGACCAGTCAGGGCAACAGCAAGACTGCTCAAGCTCACCTGTAGTCCAAGCTCACAGAGGCAG  3420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361  TTTGAGACCAGTCAGGGCAACAGCAAGACTGCTCAAGCTCACCTGTAGTCCAAGCTCACAGAGGCAG  3420

Qy   3421  AATTACCTGGGTATGGTGTGCTCACCTGTAGTCCAAGCTCACCTGTAGTCCAAGCTACACAGGAGCGCAGA  3480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421  AATTACCTGGGTATGGTGTGCTCACCTGTAGTCCAAGCTCACCTGTAGTCCAAGCTACACAGGAGCGCAGA  3480

Qy   3481  AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGCTGTCATCGATGACTAAAATAACTTT  3540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481  AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGCTGTCATCGATGACTAAAATAACTTT  3540

Qy   3541  AGTCTGGGTCGCAGTGCAGAAGCGTGTCTCAAAAATAATAAAAAACACACACTAGAGAATGTTGCAAATTGA  3600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541  AGTCTGGGTCGCAGTGCAGAAGCGTGTCTCAAAAATAATAAAAAACACACACTAGAGAATGTTGCAAATTGA  3600

Qy   3601  AAAAACAAAAAATTAATTAAATTTAAAAAACACACACTAGAGCTAGAGGAACACTAGGGCTCGCCACT  3660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601  AAAAACAAAAAATTAATTAAATTTAAAAAACACACACTAGAGCTAGAGGAACACTAGGGCTCGCCACT  3660

Qy   3661  TTATTGGGAGTCTATATCCCTGGAAGTTAATTTAAAATATTAGAAGAGTTCTTCCTCA  3720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661  TTATTGGGAGTCTATATCCCTGGAAGTTAATTTAAAATATTAGAAGAGTTCTTCCTCA  3720

Qy   3721  TTTCCTAGAGACGTCGAATGTAAATATGAATTGTAAATATCAGAGGAACACTAGGGCTCGCCACT  3780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721  TTTCCTAGAGACGTCGAATGTAAATATGAATTGTAAATATCAGAGGAACACTAGGGCTCGCCACT  3780

Qy   3781  CCAAAGTGTGTCCAAGGACCATCGCAGCAGCATCGGGAACGTAACCTGGGAACGTTAGAAATGCAG  3840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781  CCAAAGTGTGTCCAAGGACCATCGCAGCAGCATCGGGAACGTAACCTGGGAACGTTAGAAATGCAG  3840

Qy   3841  AGTCTTAGGCCTCACCCCAGACCTACTGCATTAACAGATTCTAGGT  3900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841  AGTCTTAGGCCTCACCCCAGACCTACTGCATTAACAGATTTCTAGGT  3900

Qy   3901  GCCTCACGGGCACATTAAAACTTAAAAATCCTCACTCACCACCCTTTC  3960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901  GCCTCACGGGCACATTAAAACTTAAAAATCTGCACTAGAAATCTTCACTCCACCTTTC  3960

Qy   3961  ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAAATTGATTATTTTTAATTTTCAGAA  4020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961  ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAAATTGATTATTTTTAATTTCAGAA  4020

Qy   4021  CCTTCTATTTAGGTCATCATCATATTGCTAATAGCAGGGAAGAAAGCCAAACTCTTTAACT  4080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021  CCTTCTATTTAGGTCATCATCATATTGCTAATAGCAGGGAAGAAAGCCAAACTCTTTAACT  4080

Qy   4081  GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG  4140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081  GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG  4140

Qy   4141  TGGAGCAAGTCTGTTTGATTTGGCTCGGGGCTCAGGCGCCGGCCTGTTGTGTGAATTTCACAATT  4200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141  TGGAGCAAGTCTGTTTGATTTGGCTCGGGGCTCAGGCGCCGGCCTGTTGTGTGAATTTCACAATT  4200

Qy   4201  CACAGATGTTAGCCGCTCTCGGGCTAAGTAAGTAAAGGAAGGAAGAGAATGTCAAGTTTTAAATAGCT  4260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201  CACAGATGTTAGCCGCTCTCGGGCTAAGTAAGTAAAGGAAGGAAGAGAATGTCAAGTTTAAATAGCT  4260

Qy   4261  TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTATGAAACACATTTTGAGT  4320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261  TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTATGAAACACATTTTGAGT  4320

Qy   4321  TAGATTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCTCACAAC  4380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321  TAGATTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC  4380

Qy   4381  TTTGACATCTACTGATGTCACCATGTTCACCATGTGTCCTGCTGACTAGGGGTGAAGGGA  4440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381  TTTGACATCTACTGATGTCACCATGTTCACCATGTGTCCTGCTGACTAGGGGTGAAGGGA  4440

Qy   4441  AGATGTGAACTGTAGTGTCAGTGAGCTTGACCGTTAGATACACAGAGTGGTTTTTTCCCCCTG  4500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441  AGATGTGAACTGTAGTGTCAGTGAGCTTGACCGTTAGATACACAGAGTGGTTTTTTTCCCCCTG  4500

Qy   4501  TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAATCATATTTCAAATCCACAAA  4560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501  TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAATCATATTTCAAATCCACAAA  4560

Qy   4561  ACCAGGATAAGTTTACAGCCCATATTCAGAAGGAAATAATATTTGTGTAGACTT  4620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561  ACCAGGATAAGTTTACAGCCCATATTCAGAAGGAAATAATATTTGTGTAGACTT  4620

Qy   4621  TCCTGATATTACACTGATTTGGAATATGCAAAACAGCAAAAAACTGTGAAGACATAAAGAATAGAGTGGAG  4680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621  TCCTGATATTACACTGATTTGGAATATGCAAAACAGCAAAAAACTGTAAGACATAAAGAATAGAGTGGAG  4680

Qy   4681  GTGCACTTCAGAAATATTCTACAATATAATTCCAATTTAATATCTTAAGAAAAA  4740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681  GTGCACTTCAGAAATATTCTACAATATAATTCCAATTTAATATCTTAAGAAAAA  4740

Qy   4741  CCGACTGAGAGATTAAAAATAAACTAGAATATTTTATTACAACGGCAATTTGAAAATAATTT  4800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741  CCGACTGAGAGATTAAAAATAAACTAGAATATTTTATTACAACGGCAATTTGAAAATAATTT  4800

Qy   4801  CTTTCCTTAGCTGAGGAGAGTGAGTTCAAGTTAATTAATATCTTAATATCTTAACTCAATAA  4860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801  CTTTCTCTTAGCTGAGGAGAGTGAGTTCTTAAGTTAATTAATATCTTAATATCTTAACTCAATAA  4860

Qy   4861  TTACTATATTATATGTAAGTACATGTGCATGTGTTTTGAGGTTGAGGTTAGGTAGGTATATTTAACTCAATAA  4920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861  TTACTATATTATATGTAAGTACATGTGCATGTGTTTTGAGGTTGAGGTTAGGTAGGTATATTTAACTCAATAA  4920

Qy   4921  AGGTTATTTTCTTTATTCGGGTGAGGAGAGTGAGTTCTCAAGGAGGGATGTGAAAGGGATATCT  4980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921  AGGTTATTTTCTTTATTCGGGTGAGGAGAGTGAGTTCTCAAGGAGGGATGTGAAAGGGATATCT  4980

Qy   4981  CTTTCTTAGCTGAGGAGAGTGAGTTAATAATATAATCAAGGAATTCCCT  5040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981  CTTTCTCTTAGCTGAGGAGAGTGAGTTAGTTCTTAAGTTAATATAATCAAGGAATTCCCT  5040

Qy   5041  GTCTTTGCTATTTGAGATTTGACCACCAACAGGGCCGGTTGGCTGAAAGGGAAAACTGAAGGG  5100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041  GTCTTTGCTATTTGAGATTTGACCACCAACAGGGCCGGTTGGCTGAAAGGGAAAACTGAAGGG  5100

Qy   5101  CGGGGAGGGAGGAAATAGATGAAATAAACAAAACAAAAACTTCCCTAAGCAGCTCT  5160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
ORIGIN

        Query Match          100.0%;   Score 5673.4;   DB 6;   Length 13940;
        Best Local Similarity 100.0%;   Pred. No. 0;
        Matches 5674;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   0;

Qy           2 AGCTGTTCAGGGATGTCTGGAGCACCCACCCACCATTGCTTCTGGACACTGGGTGTG   61
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          51 ATCTGTTCAGGGATGTCTGGAGCACCCACCCACCATTGCTTCTGGACACTGGGTGTG   110

Qy          62 ACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCAATT   121
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         111 ACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCAATT   170

Qy         122 GGCCTCGTTCCCTGCATGCCCTCTTCCTGGGACACTCCCTTAAGGCATCTTCTTGA   181
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         171 GGCCTCGTTCCCTGCATGCCCTCTTCCTGGGACACTCCCTTAAGGCATCTTCTTGA   230

Qy         182 CATTAACTTAACTATAAAGTCTTATTTGATGAATTTCAGTGACCTGAAGAGAGATGGAGG   241
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         231 CATTAACTTAACTATAAAGTCTTATTTGATGAATTTCAGTGACCTGAAGAGAGATGGAGG   290

Qy         242 TCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTCATTGAATTAAA   301
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         291 TCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTCATTGAATTAAA   350

Qy         302 GTCATTCGAATACCATTCAGTTACTTAAGTCTCTAGGCCCAGCTTTACTCCTAATCGATG   361
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         351 GTCATTCGAATACCATTCAGTTACTTAAGTCTCTAGGCCCAGCTTTACTCCTAATCGATG   410

Qy         362 TCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGAC   421
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         411 TCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGAC   470

Qy         422 AGAACTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTAC   481
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         471 AGAACTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTAC   530

Qy         482 AGAAAAGTGTCATACACATTAAAAATCACCAAAACTCACCAAACTCAGTCTCTCAATTGAGCAATAGTT   541
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         531 AGAAAAGTGTCATACACATTAAAAATCACCAAAACTCACCAAACTCAGTCTCTCAATTGAGCAATAGTT   590

Qy         542 GGTGAATTACTCCACCACCTCCTCCTCCTTGAAGGTTCTTTCCTGCTCTCCTCCTCACTATAA   601
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         591 GGTGAATTACTCCACCACCTCCTCCTCCTTGAAGGTTCTTTCCTGCTCTCCTCCTCACTATAA   650

Qy         602 ATGCAGGATGACCTGAAAGGCTAGGACCTGAGGTTCAGTTACCCGTACCCTGACACAAGGAATT   661
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         651 ATGCAGGATGACCTGAAAGGCTAGGACCTGAGGTTCAGTTACCCGTACCCTGACACAAGGAATT   710

Qy         662 CAGTTCTCTGATCTCATAGTCACAGTGCAGCGTGCCAGGACAATGCAAGATCAGCAAGATCATC   721
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         711 CAGTTCTCTGATCTCATAGTCACAGTGCAGCGTGCCAGGACAATGCAAGATCAGCAAGATCATC   770

Qy         722 TGCTTTAAGCCTCTTGTGTGGCATCGTGTTTTCCACTGCGCCTGTACCTATTGCTCTT   781
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         771 TGCTTTAAGCCTCTTGTGTGGCATCGTGTTTTCCACTGCGCCTGTACCTATTGCTCTT   830

Qy         782 TCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCGTCCAGTCATGGTAGGGC   841
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         831 TCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCGTCCAGTCATGGTAGGGC   890

Qy         842 CATCAGTCCACATGATCAGGCCTCTCCTGGCCAAACATGGCATCTTCTTTTTGGAATTT   901
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         891 CATCAGTCCACATGATCAGGCCTCTCCTGGCCAAACATGGCATCTTCTTTTTGGAATTT   950

Qy         902 GAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCGACTTACGCCTACA   961
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         951 GAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCGACTTACGCCTACA   1010

Qy         962 GTGGCTTTGCAAAGTGACTGCTGTCCATTCCTATTTCTTAAGTCCCGAATTTATAATTTATC   1021
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1011 GTGGCTTTGCAAAGTGACTGCTGTCCATTCCTATTTCTTAAGTCCCGAATTTATAATTTATC   1070
```

```
Db        5101 CGGGGAGGAGGGAAATAGATGAAAAAACAAAACAAAACAAAACTTCCCTAAGCAGCTCT   5160

Qy        5161 ACAAAACATTTTAGCCCAGAAATAGTCACAGAGAATCCTCAAATCAAACCAGTATCCAGA   5220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5161 ACAAAACATTTTAGCCCAGAAATAGTCACAGAGAATCCTCAAATCAAACCAGTATCCAGA   5220

Qy        5221 TACAAGGAAGTGTTATGTGAGCGAGCAGGGTGGACACTCATCAGCTCAGTTCAGTTACA   5280
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5221 TACAAGGAAGTGTTATGTGAGCGAGCAGGGTGGACACTCATCAGCTCAGTTCAGTTACA   5280

Qy        5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCAGCATCAATGCCAGCATCCAATCACGACA   5340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCAGCATCAATGCCAGCATCCAATCACGACA   5340

Qy        5341 GAGATCAGAAGTTCAGAGACTTCAGACGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTA   5400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5341 GAGATCAGAAGTTCAGAGACTTCAGACGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTA   5400

Qy        5401 TACGTCAAGGACTCTGAAGCCGTGAGAGGGGAAGAACAACAGTAGAGAGAGGATGCCCA   5460
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5401 TACGTCAAGGACTCTGAAGCCGTGAGAGGGGAAGAACAACAGTAGAGAGAGGATGCCCA   5460

Qy        5461 GCTGTAAGAATCAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA   5520
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5461 GCTGTAAGAATCAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA   5520

Qy        5521 TCAAGAAACGCTCCGCCCTATTGCTTTTGCAAATATGTGAAGGAGAGAAGTGCCTAAACTTCTA   5580
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5521 TCAAGAAACGCTCCGCCCTATTGCTTTTGCAAATATGTGAAGGAGAGAAGTGCCTAAACTTCTA   5580

Qy        5581 TGTCTGATAGCATTTGACCCTATTGCTTTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTCTATATACAC   5640
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5581 TGTCTGATAGCATTTGACCCTATTGCTTTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTCTATATACAC   5640

Qy        5641 AGGTATTTGTGTATATTTTATATATTTTATATATAATTGTTCTCCGT   5676
               |||||||||||||||||||||||||||||||||||||||||||||||
Db        5641 AGGTATTTGTGTATATTTTATATATTTTATATATAATTGTTCTCCGT   5676
```

```
RESULT 5
BD093100
LOCUS       BD093100       13940 bp   DNA   linear   PAT   27-AUG-2002
DEFINITION  Gene encoding promoter region of tumor suppressor gene p51.
ACCESSION   BD093100
VERSION     BD093100.1   GI:22638688
KEYWORDS    WO 0100818-A/12.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 13940)
  AUTHORS   Sakai,T., Kagaya,S., Sato,T., Sukenaga,Y. and Fujii.H.
  TITLE     Gene encoding promoter region of tumor suppressor gene p51
  JOURNAL   Patent: WO 0100818-A 12 04-JAN-2001;
            TOSHIYUKI SAKAI,NIPPON KAYAKU KK,SHIGEHIDE KAGAYA,TAKAMICHI SATO,
            YOSHIKAZU SUKENAGA,HIDEJI FUJII
  OS        Artificial Sequence
  PN        WO 0100818-A/12
  PD        04-JAN-2001
  PF        28-JUN-2000 WO 2000JP004261
  PR        29-JUN-1999 JP 99P  183195
  PI        TOSHIYUKI SAKAI,SHIGEHIDE KAGAYA,TAKAMICHI SATO,YOSHIKAZU  PI
            SUKENAGA,
  PI        HIDEJI FUJII
  PC        C12N15/10,C12N5/10,C12Q1/68,A61K48/00,A61K45/00,A61P43/00, PC
            A61P35/00
COMMENT     Plasmid gene containing p51promoter and neomycin resistance
            gene
FEATURES             Location/Qualifiers
     source          1..13940
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
```

```
Qy  1022  CTGGTTACAGCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTTATAGTCTGTG  1081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1071  CTGGTTACAGCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTTATAGTCTGTG  1130

Qy  1082  AATTTTCATATTTCTTTTCATACATTTCATGTTTTGTTTGTTGTTGTGTGTTTTTTTG  1141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131  AATTTTCATATTTCTTTTCATACATTTCATGTTTTGTTTGTTGTTGTGTGTTTTTTTG  1190

Qy  1142  GCTTTAGGTAGGCAGAATCAGTTCTGTTGTTATACCCAAGGAATCCTGATTGATACAT  1201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191  GCTTTAGGTAGGCAGAATCAGTTCTGTTGTTATACCCAAGGAATCCTGATTGATACAT  1250

Qy  1202  CCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGGT  1261
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1251  CCTTCCCCTTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGGT  1310

Qy  1262  CTGGGAGTAAGTAAGTACCAGAGCTCCGTACTAACCCCAGGTTAGCCAACTGCTTTACACA  1321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1311  CTGGGAGTAAGTAAGTACCAGAGCTCCGTACTAACCCCAGGTTAGCCAACTGCTTTACACA  1370

Qy  1322  ACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAAGAAGCTACTATTTTGCC  1381
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1371  ACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAAGAAGCTACTATTTTGCC  1430

Qy  1382  AAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAAA  1441
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1431  AAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAAA  1490

Qy  1442  GTTGTTGGACTTCTGGCCATAATGTGTATCTAAGACCAGATTCATTCTTAATAGCTA  1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1491  GTTGTTGGACTTCTGGCCATAATGTGTATCTAAGACCAGATTCATTCTTAATAGCTA  1550

Qy  1502  AACAAACAAAGAAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGTAACA  1561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1551  AACAAACAAAGAAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGTAACA  1610

Qy  1562  GTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGACA  1621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1611  GTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGACA  1670

Qy  1622  CAGAGAATACGTACCCATATGCCTTGCATTTATGTGTATTCTAGACATTGTATTGTCCAAC  1681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1671  CAGAGAATACGTACCCATATGCCTTGCATTTATGTGTATTCTAGACATTGTATTGTCCAAC  1730

Qy  1682  ATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTGAAATTAAGTAAGTGAATTAAGAA  1741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1731  ATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTGAAATTAAGTAAGTGAATTAAGAA  1790

Qy  1742  TAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAACTAGTGCTACCACATCAGACG  1801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1791  TAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAACTAGTGCTACCACATCAGACG  1850

Qy  1802  GTGCAAATATAGAATATATTTCCTTTATAACAGAAGTTCTATTGGAAAACAATGTTCTAGA  1861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1851  GTGCAAATATAGAATATATTTCCTTTATAACAGAAGTTCTATTGGAAAACAATGTTCTAGA  1910

Qy  1862  AAATATACACATAATCTATAAAAACAAAAAGCAAGTCAGTTGCTCTAAGGCCAGGGGT  1921
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1911  AAATATACACATAATCTATAAAAACAAAAAGCAAGTCAGTTGCTCTAAGGCCAGGGGT  1970

Qy  1922  GAGGGGAGATCGATTGCAAAGTGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGAA  1981
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1971  GAGGGGAGATCGATTGCAAAGTGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGAA  2030

Qy  1982  TCTTGATTGCCGATGAAGGCTACTCGGTGTCCGGTGTCTAATGTGTCACCTCCTGCAGACTACTT  2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2031  TCTTGATTGCCGATGAAGGCTACTCGGTGTCCGGTGTCTAATGTGTCACCTCCTGCAGACTACTT  2090

Qy  2042  GGAATTGGCGAATTTCATTGTATGTAATTATACCTCATAAAGTAACTCTAAGAGGTCAA  2101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2091  GGAATTGGCGAATTTCATTGTATGTAATTATACCTCATAAAGTAACTCTAAGAGGTCAA  2150
```

```
Qy  2102  GTGTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCAA  2161
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2151  GTGTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCAA  2210

Qy  2162  ATACATAAACATGTTATTCATTCCATTAGGTGCAATATTTTTGCTAGCTCCTGAAAACACA  2221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2211  ATACATAAACATGTTATTCATTCCATTAGGTGCAATATTTTTGCTAGCTCCTGAAAACACA  2270

Qy  2222  GAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCGTTCACAGGAGGATGA  2281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2271  GAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCGTTCACAGGAGGATGA  2330

Qy  2282  GTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAAAG  2341
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2331  GTCTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAAAG  2390

Qy  2342  TGCAGATATAATTATATGTAGGACAGTAGAAGTGGGGAGGTTTCTTTTATGTGGAAAAAAG  2401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2391  TGCAGATATAATTATATGTAGGACAGTAGAAGTGGGGAGGTTTCTTTTATGTGGAAAAAAG  2450

Qy  2402  AGGGAGGAATTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGTGAGTTTT  2461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2451  AGGGAGGAATTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGTGAGTTTT  2510

Qy  2462  AAAACATTCCTGGTGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATAT  2521
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2511  AAAACATTCCTGGTGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATAT  2570

Qy  2522  GCAACCTAGAGGAAAGTGCATGAAGGGAGCAGTTGTAAAATAATTTCATGAATGTAAG  2581
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2571  GCAACCTAGAGGAAAGTGCATGAAGGGAGCAGTTGTAAAATAATTTCATGAATGTAAG  2630

Qy  2582  TGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCTCCTAG  2641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2631  TGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCTCCTAG  2690

Qy  2642  GAGTCAAGAAGAACAAGTGTCCCTTTCTCCTCACGTTATGCTCAGTTGGTCCAAGTCCAAA  2701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2691  GAGTCAAGAAGAACAAGTGTCCCTTTCTCCTCACGTTATGCTCAGTTGGTCCAAGTCCAAA  2750

Qy  2702  ACACCTTCCTTCCTTAAGTACTTTCTTCTTCTCCCCATCACAAATCTAAAGTCTTCACAA  2761
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2751  ACACCTTCCTTCCTTAAGTACTTTCTTCTTCTCCCCATCACAAATCTAAAGTCTTCACAA  2810

Qy  2762  ACATCATTCATTTAAACAGGCAGTCATGTCAGAAGGCAATTGCTTTTCCTAGACTTCTATG  2821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2811  ACATCATTCATTTAAACAGGCAGTCATGTCAGAAGGCAATTGCTTTTCCTAGACTTCTATG  2870

Qy  2822  TACGTTATTATTACAATTCTGCCTAAAGTCTTGGTGTTCTGGAAAGTTCCACC  2881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2871  TACGTTATTATTACAATTCTGCCTAAAGTCTTGGTGTTCTGGAAAGTTCCACC  2930

Qy  2882  TTGCACATCAAAGATATAAATTCATGCATTGTATAGTAACCTTGTATAGTAACCTTAGTCCCCTAAGAGAATA  2941
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2931  TTGCACATCAAAGATATAAATTCATGCATTGTATAGTAACCTTGTATAGTAACCTTAGTCCCCTAAGAGAATA  2990

Qy  2942  AGGATGAACTATATAAATATAAGAGTAATTGGTAATTATAATGGTAATTATATATATTT  3001
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2991  AGGATGAACTATATAAATATAAGAGTAATTGGTAATTATAATGGTAATTATATATATTT  3050

Qy  3002  TTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTTCTGTTGAATTCTAGAGAGTTTGC  3061
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3051  TTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTTCTGTTGAATTCTAGAGAGTTTGC  3110

Qy  3062  CTCTTTTTCCTGGGTCAACTCTCGCCATTTATTTCCATAATGCAATAGGAGCCAATCTTT  3121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3111  CTCTTTTTCCTGGGTCAACTCTCGCCATTTATTTCCATAATGCAATAGGAGCCAATCTTT  3170

Qy  3122  TTCATAATTACTTATTTAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTAA  3181
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3171  TTCATAATTACTTATTTAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTAA  3230

Qy  3182  CTTTAGGATTTTTAAATAACAACTATTGAAAATCATGACATACGTTTAAATGATATTATTT  3241
```

Db 4311 CTCCCTTCCATCCTGCTGAAGCAACAAATAAAATATTTTTATGAAACACATTTTGAGTT 4370
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 4322 AGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGTTGTCTCACAACT 4381
Db 4371 AGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGTTGTCTCACAACT 4430
Qy 4382 TTGACATCTACTGATGTCACCTATTTACACGGTGTGTCCTGTGACTAGGGGTGAAGGGAA 4441
Db 4431 TTGACATCTACTGATGTCACCTATTTACACGGTGTGTCCTGTGACTAGGGGTGAAGGGAA 4490
Qy 4442 GATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTCCCCTGT 4501
Db 4491 GATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTCCCCTGT 4550
Qy 4502 TGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCAAATCCACAAAA 4561
Db 4551 TGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCAAATCCACAAAA 4610
Qy 4562 CCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATTTATTTGTGTGTGAGACTTT 4621
Db 4611 CCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATTTATTTGTGTGTGAGACTTT 4670
Qy 4622 CCTGATATTACACTGATTTGGGAATATATGGAATAATGGAACAATTTATGGTTTCCTTTCGAAGTAGG 4681
Db 4671 CCTGATATTACACTGATTTGGGAATATATGGAATAATGGAACAATTTATGGTTTCCTTTCGAAGTAGG 4730
Qy 4682 TCAAGTCAAAGCAAAACCAAAAACAGCAAAAACAGTGTAAGACATAAAGAAGTGGAGC 4741
Db 4731 TCAAGTCAAAGCAAAACCAAAAACAGCAAAAACAGTGTAAGACATAAAGAAGTGGAGC 4790
Qy 4742 CGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATATTTG 4801
Db 4791 CGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATATTTG 4850
Qy 4802 TGCACTTCAGAATATTCTACAATAATATATATTTAATATCTTTAAGAAAAT 4861
Db 4851 TGCACTTCAGAATATTCTACAATAATATATATTTAATATCTTTAAGAAAAT 4910
Qy 4862 TACTATATTATGTAAGTACATGTCATGTGTTTGAGGTAGGATATTTAAACTCAATAAA 4921
Db 4911 TACTATATTATGTAAGTACATGTCATGTGTTTGAGGTAGGATATTTAAACTCAATAAA 4970
Qy 4922 GGTTATTTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCTC 4981
Db 4971 GGTTATTTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCTC 5030
Qy 4982 TTTCTCTTAGCTGAGAGGAAGGAGTGAGTTCTAAGTTAAATAATCAAGGAATTTCCCTG 5041
Db 5031 TTTCTCTTAGCTGAGAGGAAGGAGTGAGTTCTAAGTTAAATAATCAAGGAATTTCCCTG 5090
Qy 5042 TCTTTGCTATTTGAGATTGTGACCCACACAGCCGGTTGGCTGAAAGGGAAACTGAAGGGC 5101
Db 5091 TCTTTGCTATTTGAGATTGTGACCCACACAGCCGGTTGGCTGAAAGGGAAACTGAAGGGC 5150
Qy 5102 GGGGAGGAGGGAAATAGATGAAATAGATGAAAAACAAAACAAAACTTCCCTAAGCAGCTCTA 5161
Db 5151 GGGGAGGAGGGAAATAGATGAAATAGATGAAAAACAAAACAAAACTTCCCTAAGCAGCTCTA 5210
Qy 5162 CAAAACATTTTAGCCCCAGAAAATAGTCACAGAAATCCTCAAATCAAACCAGTCAGATCAGAT 5221
Db 5211 CAAAACATTTTAGCCCCAGAAAATAGTCACAGAAATCCTCAAATCAAACCAGTCAGATCAGAT 5270
Qy 5222 ACAAGGAAGTGTTATGTGAGCTGGAGCAGGGTGGACACTCATCAGCTTCAGTTGACTACAA 5281
Db 5271 ACAAGGAAGTGTTATGTGAGCTGGAGCAGGGTGGACACTCATCAGCTTCAGTTGACTACAA 5330
Qy 5282 AAGTCCAGGCTGCTCAGATCCAATTCATGCCAGATCACGACAGCTCTA 5341
Db 5331 AAGTCCAGGCTGCTCAGATCCAATTCATGCCAGATCACGACAGCTCTA 5390
Qy 5342 AGATCAGAAGTTCAGAGAGATGCCTCCAAATTGCCAACAACAAGTGTGGCTACTAT 5401
Db 5391 AGATCAGAAGTTCAGAGAGATGCCTCCAAATTGCCAACAACAAGTGTGGCTACTAT 5450

Qy 3231 CTTTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTAAATGATATTATTT 3290
Db 3242 AAATACGTTAGGCTATAAACCTTTAAATTTTTAAATTTTTAAAATTTTTAAAATTTTTAAATTTTT 3301
Qy 3291 AAATACGTTAGGCTATAAACCTTTAAATTTTTAAATTTTTAAAATTTTTAAATTTTTGGTGGCTC 3350
Db 3302 ATGCCTGTAATCCCAACACACTTTGGGAGGATAGCTTGAGTCCAGCAGT 3361
Qy 3351 ATGCCTGTAATCCCAACACACTTTGGGAGGATAGCTTGAGTCCAGCAGT 3410
Db 3362 TTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAACAAAAACAAA 3421
Qy 3411 TTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAACAAAAACAAA 3470
Db 3422 ATTACCTGGGTATGGTGTCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAA 3481
Qy 3471 ATTACCTGGGTATGGTGTCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAA 3530
Db 3482 GGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACTCA 3541
Qy 3531 GGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACTCA 3590
Db 3542 GTCTGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAAATAAAATAACTTTTA 3601
Qy 3591 GTCTGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAAATAAAATAACTTTTA 3650
Db 3602 AAAAACAAAAATTAATTAAATTTTTAAAACACAACACTAGAGATGTTTGCAAATTGAT 3661
Qy 3651 AAAAACAAAAATTAATTAAATTTTTAAAACACAACACTAGAGATGTTTGCAAATTGAT 3710
Db 3662 TATTTGGGAGTCTATATCCCTGGAGTTAATTTAAAATATTTAGAGAGTTCTTCCTCAT 3721
Qy 3711 TATTTGGGAGTCTATATCCCTGGAGTTAATTTAAAATATTTAGAGAGTTCTTCCTCAT 3770
Db 3722 TTCCTAGAGACGTCGAATTGTAAATATCAGAGCTAGAAGGAACACTAGGGCTCGCCACTC 3781
Qy 3771 TTCCTAGAGACGTCGAATTGTAAATATCAGAGCTAGAAGGAACACTAGGGCTCGCCACTC 3830
Db 3782 CAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACCGTGTTAGAAATGCAGA 3841
Qy 3831 CAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACCGTGTTAGAAATGCAGA 3890
Db 3842 GTCTTAGGCTCACCCCAGCACCTGACCAGAATCTGCATTAACAGATTTCTAGGTG 3901
Qy 3891 GTCTTAGGCTCACCCCAGCACCTGACCAGAATCTGCATTAACAGATTTCTAGGTG 3950
Db 3902 CCTCACGGGCACATTAGTAATAATTAATTAGTGAAGCAATCTTCCCCTTCACTCCACCTTCA 3961
Qy 3951 CCTCACGGGCACATTAGTAATAATTAATTAGTGAAGCAATCTTCCCCTTCACTCCACCTTCA 4010
Db 3962 TTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATTGATTATTTTAAATTCACAATTC 4021
Qy 4011 TTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATTGATTATTTTAAATTCACAATTC 4070
Db 4022 CTTCTATTTAGGTCATCTATATTGCTAATAGCAGGGAAGCCAAACTCTTTAACTG 4081
Qy 4071 CTTCTATTTAGGTCATCTATATTGCTAATAGCAGGGAAGCCAAACTCTTTAACTG 4130
Db 4082 CAATTAACAAATCTATAATTAATAGTTAAGCAATCTTCCCCTTTAAGTTTTACATTTTGT 4141
Qy 4131 CAATTAACAAATCTATAATTAATAGTTAAGCAATCTTCCCCTTTAAGTTTTACATTTTGT 4190
Db 4142 GGAGCAAGCTGTTTGATTTGGCTGGGCTCGGGCTCAGGGCCTGTTTGTGAAATTCACAATTC 4201
Qy 4191 GGAGCAAGCTGTTTGATTTGGCTGGGCTCGGGCTCAGGGCCTGTTTGTGAAATTCACAATTC 4250
Db 4202 ACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAGAATTGTCAAGTTTTAAATAGTCTT 4261
Qy 4251 ACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAGAATTGTCAAGTTTTAAATAGTCTT 4310
Db 4262 CTCCCTTCCATCCTGCTGAAGCAACAAATAAAATATTTTTATGAAACACATTTTGAGTT 4321

```
Qy  5402  ACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGGAGAACAACAGTAGAGAGGATGCCCAG  5461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5451  ACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGGAGAACAACAGTAGAGAGGATGCCCAG  5510

Qy  5462  CTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAAT  5521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5511  CTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAAT  5570

Qy  5522  CAAGAAACGCTCCGCCTCTTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTAT  5581
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5571  CAAGAAACGCTCCGCCTCTTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTAT  5630

Qy  5582  GTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATATACACA  5641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5631  GTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATATACACA  5690

Qy  5642  GGTATTTGTGTATATTTTATATAATTGTTCTCCGT  5676
          |||||||||||||||||||||||||||||||||||
Db  5691  GGTATTTGTGTATATTTTATATAATTGTTCTCCGT  5725
```

```
RESULT 6
AC078809/c
LOCUS       AC078809     162646 bp    DNA     linear   PRI 25-JAN-2003
DEFINITION  Homo sapiens 3 BAC RP11-600G3 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC078809
VERSION     AC078809.9  GI:19033390
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162646)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,D., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newton,J., Newson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,K., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162646)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 162646)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 162646)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 162646)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 162646)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 1, 2002 this sequence version replaced gi:17977513.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.
```

```
FEATURES        QUALSTAT-REPORT.
source          Location/Qualifiers
                1..162646
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="RP11-600G3"
misc_feature    1..2004
                /note="overlaps bases 79511..81510 of clone AC117486"
                /function="clone overlap"
repeat_region   complement(248..339)
                /rpt_family="MIR"
repeat_region   491..556
                /rpt_family="(TA)n"
repeat_region   complement(844..981)
                /rpt_family="THE1B"
repeat_region   1586..1727
                /rpt_family="MIR"
repeat_region   1683..1736
                /rpt_family="L2"
repeat_region   2734..2922
                /rpt_family="MIR"
repeat_region   2977..3083
                /rpt_family="L2"
repeat_region   complement(3383..3562)
                /rpt_family="MIR"
repeat_region   3742..7012
                /rpt_family="L1MC3"
repeat_region   7013..7275
                /rpt_family="AluSx"
repeat_region   7276..7892
                /rpt_family="L1MC3"
repeat_region   complement(8146..8365)
                /rpt_family="MIR"
repeat_region   9447..9479
                /rpt_family="MIR"
repeat_region   9648..9780
                /rpt_family="AT_rich"
repeat_region   10440..10660
                /rpt_family="GA-rich"
repeat_region   complement(10703..11014)
                /rpt_family="MER20"
repeat_region   11431..11683
                /rpt_family="AluSx"
STS             /standard_name="79189"
                11911..12013
repeat_region   complement(11911..12013)
                /rpt_family="MIR"
repeat_region   12654..12964
                /rpt_family="AluSx"
repeat_region   13022..13142
                /rpt_family="MIR"
repeat_region   15041..15337
                /rpt_family="AluSg"
repeat_region   15456..15480
                /rpt_family="(TTTA)n"
repeat_region   complement(16012..16151)
                /rpt_family="MIR"
repeat_region   17138..17449
                /rpt_family="L2"
repeat_region   17765..17895
                /rpt_family="AluSp"
repeat_region   18694..18860
                /rpt_family="MIR"
repeat_region   18907..19141
                /rpt_family="MIR"
repeat_region   19474..19515
                /rpt_family="(TA)n"
repeat_region   complement(19940..20118)
                /rpt_family="AluSx"
repeat_region   20726..21393
                /rpt_family="L1M3e"
repeat_region   complement(21402..21704)
                /rpt_family="L1"
```

```
repeat_region   .21710..23051
                /rpt_family="L1M2"
repeat_region   23048..24606
                /rpt_family="L1M1"
repeat_region   24611..24765

Query Match            99.3%;    Score 5635.6;   DB 9;   Length 162646;
Best Local Similarity  99.8%;    Pred. No. 0;
Matches 5666;  Conservative  0;  Mismatches  9;  Indels  5;  Gaps  2;

Qy       1 CAGCTGTTCAGGGATGTCTGGAAAAGAAGGCCACCCACATTGCTTCTGGACACTGGGTGT     60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76260 CAGCTGTTCAGGGATGTCTGGAAAAGAAGGCCACCCACATTGCTTCTGGACACTGGGTGT  76201

Qy      61 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT    120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76200 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT   76141

Qy     121 TGGCCCTCGTTCCCTGCATGCCCTCTTTCCTTGGGACACTCCCTAAGGCATCTTCTTG    180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76140 TGGCCCTCGTTCCCTGCATGCCCTCTTTCCTTGCGGACACTCCCTAAGGCATCTTCTTG   76081

Qy     181 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACTGAAGAGAGATGGAG    240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76080 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACTGAAGAGAGATGGAG   76021

Qy     241 GTCAAATCAGAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA    300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76020 GTCAAATCAGAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA   75961

Qy     301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTCAGGCCCAGCTTACTCCTAATCGAT    360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75960 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTCAGGCCCAGCTTACTCCTAATCGAT   75901

Qy     361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA    420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75900 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA   75841

Qy     421 CAGAACTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA    480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75840 CAGAACTTGGCTTCCACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA   75781

Qy     481 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTCTTCAATTTGAGCAATAGT    540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75780 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTCTTCAATTTGAGCAATAGT   75721

Qy     541 TGGTGAATTACTCACCACCTCCTCCTGAAGGTTCTTTCCTGCTCTCCTCACTATA    600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75720 TGGTGAATTACTCACCACCTCCTCCTGAAGGTTCTTTCCTGCTCTCCTCACTATA   75661

Qy     601 AATGCAGGATGACCTGGAAAGCTAGGACACTGAGGTTCAGTTACCCTGACACAAAGGAAT    660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75660 AATGCAGGATGACCTGGAAAGCTAGGACACTGAGGTTCAGTTACCCTGACACAAAGGAAT   75601

Qy     661 TCAGTTTCTCTGATCTCATAGTCACAGGCTGCCAGAGCTCTACGGAACATGCAAGATCAT    720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75600 TCAGTTTCTCTGATCTCATAGTCACAGGCTGCCAGAGCTCTACGGAACATGCAAGATCAT   75541

Qy     721 CTGCTTTAAGCCTCTTGTGGTGGCATCGTTGTTTCCACTGCCCTGTACCTATTGCTCT    780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75540 CTGCTTTAAGCCTCTTGTGGTGGCATCGTTGTTTCCACTGCCCTGTACCTATTGCTCT   75481

Qy     781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCTGCTCAGTCATGGTAGGG    840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75480 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCTGCTCAGTCATGGTAGGG   75421

Qy     841 CCATCAGTCCACATGATCAGGCCTCTCCTGGCCAAACATGGCATCTTTCTTTGGGAATT    900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75420 CCATCAGTCCAGATGATCAGGCCTCTCCTGGCCAAACATGGCATCTTTCTTTGGGAATT   75361

Qy     901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGTGTTGAATCTGACTTACGCCTAC    960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75360 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAAGTGTTGAATCTGACTTACGCCTAC   75301
```

```
Qy     961 AGTGGCTTTGCAAAGTGACTGTCCATTCCTATTCTTAAGTCCCTGAATTTATAATTTAT 1020
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   75300 AGTGGCTTTGCAAAGTGACTGTCCATTCCTATTCTTAAGTCCCTGAATTTATAATTTAT 75241

Qy    1021 CCTGGTTACAGCGCCCTTCTGAGATGTGTGGTTTTTTTTCCAACTGTCTCTTATAGTCTGT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75240 CCTGGTTACAGCGCCCTTCTGAGATGTGTGGTTTTTTTTCCAACTGTCTCTTATAGTCTGT 75181

Qy    1081 GAATTTCATATTTCTTTCATACATTTCATGT----TTTGTTGTTGTTGTTGTGTGT 1136
           ||||||||||||||||||||||||||||      |||||||||||||||||||||||
Db   75180 GAATTTCATATTTCTTTCATACATTTCATGTTTGTTGTTGTTGTTGTTGTTGTGTGT 75121

Qy    1137 TTTTGGCTTTAGGTAGGCAGAATCAGTTTCGTTGTTGTTATACCCAAGGAATCCTGATTGA 1196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75120 TTTTGGCTTTAGGTAGGCAGAATCAGTTTCGTTGTTGTTATACCCAAGGAATCCTGATTGA 75061

Qy    1197 TACATCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCT 1256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75060 TACATCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCT 75001

Qy    1257 GAGGTCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAAACCCCAGGTTAGCCAACTGCTTT 1316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75000 GAGGTCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAAACCCCAGGTTAGCCAACTGCTTT 74941

Qy    1317 ACACAACATTTGCTCTCCTTCAGAGTTATAGACAGTCTTGGAAGAAGAAGCTACTATT 1376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74940 ACACAACATTTGCTCTCCTTCAGAGTTATAGACAGTCTTGGAAGAAGAAGCTACTATT 74881

Qy    1377 TTGCCAAAGACCTCAGGAGGACCAAGAACAAGTTCTCGGGATATGTGATGATTGAACTCTT 1436
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74880 TTGCCAAAGACCTCAGGAGGACCAAGAACAAGTTCTCGGGATATGTGATGATTGAACTCTT 74821

Qy    1437 AAAAGTTGTTGGACTTCTGGCCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAAT 1496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74820 AAAAGTTGTTGGACTTCTGGCCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAAT 74761

Qy    1497 AGCTAAACAAACAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTG 1556
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74760 AGCTAAACAAACAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTG 74701

Qy    1557 ATACAGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAAATCC 1616
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74700 ATACAGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAAATCC 74641

Qy    1617 AGACACAGAAGAATACGTACCATATGCCTGCATTATGTGATATTCTAGACATTGTATTGT 1676
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74640 AGACACAGAAGAATACGTACCATATGCCTGCATTATGTGATATTCTAGACATTGTATTGT 74581

Qy    1677 CCAACATAGTAGCCAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTAAATTAAGT 1736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74580 CCAACATAGTAGCCAGTAGCCAGTAGCCTCACATGGCTATTCAAATTTGATTAAATTAAGT 74521

Qy    1737 AAGAATAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAGCCACATGGCTACCACATC 1796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74520 AAGAATAAAAATTTAGCTCTTCAGTAGCGTTAGCCACATGTAGCCACATGGCTACCACATC 74461

Qy    1797 AGACCGGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTT 1856
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74460 AGACCGGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTT 74401

Qy    1857 CTAGAAAATATACACATAATCTATAAAACAAAAAGCAAGTCAGTAGTTGTCTAAGGCCA 1916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74400 CTAGAAAATATACACATAATCTATAAAACAAAAAGCAAGTCAGTAGTTGTCTAAGGCCA 74341

Qy    1917 GGGGTGAGGGGAGATCGATTGCAAAGTGCTATGGTAATGGCTGCAAAGTTTTGGGGTTGT 1976
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74340 GGGGTGAGGGGAGATCGATTGCAAAGTGCTATGGTAATGGCTGCAAAGTTTTGGGGTTGT 74281

Qy    1977 TGGAATCTTGATTGCCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAA 2036
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74280 TGGAATCTTGATTGCCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAA 74221
```

```
Qy    2037 CACTTGGAATTGGCGAATTCATTGTATGTAAATATACCTCATAAAGTAACTCTAAGAG 2096
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74220 CACTTGGAATTGGCGAATTCATTGTATGTAAATATACCTCATAAAGTAACTCTAAGAG 74161

Qy    2097 GTCAAGTGTTTGTGAAATATTTTAATCAGTTGCAATACTTATTATGAGATGATTTT 2156
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74160 GTCAAGTGTTTGTGAAATATTTTAATCAGTTGCAATACTTATTATGAGATGATTTT 74101

Qy    2157 TGCAAATACATACAAACATGTTATTCATCCATTAGGTGCAATPATTTTGCTAGCTCCTGAAA 2216
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74100 TGCAAATACATACAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTCCTGAAA 74041

Qy    2217 ACACAGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGA 2276
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74040 ACACAGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGA 73981

Qy    2277 GATGAGTCTATTCAAAAATAGCTAGACTCCAGGAGCAAAGTTATAGGTGACCTTACCACAA 2336
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73980 GATGAGTCTATTCAAAAATAGCTAGACTCCAGGAGCAAAGTTATAGGTGACCTTACCACAA 73921

Qy    2337 AAAAGTGCAGATATAATTATTATGTAGGACAGTAGCAGTGGGAAGTTCTTTTATGTGGAA 2396
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73920 AAAAGTGCAGATATAATTATTATGTAGGACAGTAGCAGTGGGAAGTTCTTTTATGTGGAA 73861

Qy    2397 AAAAGAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGGCGCAGATGA 2456
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73860 AAAAGAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGGCGCAGATGA 73801

Qy    2457 GTTTTAAAACATTCCTGGTGGAGGGCAGAATATGATTCCAAGGCACAAGAGCAACCAGAAA 2516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73800 GTTTTAAAACATTCCTGGTGGAGGGCAGAATATGATTCCAAGGCACAAGAGCAACCAGAAA 73741

Qy    2517 AATATGCAACCTAGAGAAGTGCATGAAGGGAGCAGTTGTAAAATAATTTCATGAAT 2576
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73740 AATATGCAACCTAGAGAAGTGCATGAAGGGAGCAGTTGTAAAATAATTTCATGAAT 73681

Qy    2577 GTAAGTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCT 2636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73680 GTAAGTGAGAAGAATTGTATCATAGACACCTGAGTTTGGCAGAGTGCATGTTCTTGGCT 73621

Qy    2637 CCTAGGAGTCAAGAGAACAAAGTGTCCCTTTCTCCTCACGTTATGCTCAGTGGTCCAAGT 2696
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73620 CCTAGGAGTCAAGAGAACAAAGTGTCCCTTTCTCCTCACGTTATGCTCAGTGGTCCAAGT 73561

Qy    2697 CCAAAAACACCTTCCTTCCTAAGTACTTCTTCTTCCTCCCCCATACAAATCTAAAGTCTT 2756
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73560 CCAAAAACACCTTCCTTCCTAAGTACTTCTTCTTCCTCCCCCATACAAATCTAAAGTCTT 73501

Qy    2757 CACAAACATCATTCATTTAAACAGGCAGGTCATGTCAGAAAGGCAATTGCTTTTCCTAGACTT 2816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73500 CACAAACATCATTCATTTAAACAGGCAGGTCATGTCAGAAAGGCAATTGCTTTTCCTAGACTT 73441

Qy    2817 CTATGTACGTTATTATATTCTGCCTAAAAGACTCTAAAGTCTTGGAAAAGTTGT 2876
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73440 CTATGTACGTTATTATATTCTGCCTAAAAGACTCTAAAGTCTTGGAAAAGTTGT 73381

Qy    2877 CCACCTTGCACATCAAGAGATAATTCATGCATTGTATAGTAACCTTAGTCCCCTAAGA 2936
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73380 CCACCTTGCACATCAAGAGATAATTCATGCATTGTATAGTAACCTTAGTCCCCTAAGA 73321

Qy    2937 GAATAAGGATGAACTATAAAATATAAAGAAGTAATTATGGTAATTATAATATGCCACT 2996
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73320 GAATAAGGATGAACTATAAAATATAAAGAAGTAATTATGGTAATTATAATATGCCACT 73261

Qy    2997 TATTTTTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTTCTCGTTGAATTCTAGAGAG 3056
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73260 TATTTTTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTTCTCGTTGAATTCTAGAGAG 73201

Qy    3057 TTTGCCTCTTTTTCCTGGGTCAACTTCGCGCCATTTATTTCCATAATGCAATAGGAGCCAA 3116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73200 TTTGCCTCTTTTTCCTGGGTCAACTTCGCGCCATTTATTTCCATAATGCAATAGGAGCCAA 73141

Qy    3117 TCTTTTTCATAATTACTTATTTAAAATTGTTGCCATTAATTCTGTTCCTTAGCTT 3176
           |||||||||||||||
Db   73140
```

```
Db  72061  AATTCACAGATGTTAGCCGCTTCTCGGGCTAAGTAAGGAGAGAATGTCAAGTTTAAAT  72002

Qy  4257   AGCTTCTCCCTTCCATCCTGCTGAAGCAACAAACAAATAAAATATTTTATGAAACACATTT  4316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72001  AGCTTCTCCCTTCCATCCTGCTGAAGCAACAAACAAATAAAATATTTTATGAAACACATTT  71942

Qy  4317   GAGTTAGATTACTTACAGGAGAAATGTCAAATTTCTCTGAAAGGGCTTAGATTGTCTCA  4376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71941  GAGTTAGATTACTTACAGGAGAAATGTCAAATTTCTCTGAAAGGGCTTAGATTGTCTCA  71882

Qy  4377   CAACTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGTGACTAGGGGTGAA  4436
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71881  CAACTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGTGACTAGGGGTGAA  71822

Qy  4437   GGGAAGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTTCCC  4496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71821  GGGAAGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTTCCC  71762

Qy  4497   CCTGTTGGAGTCTATCCTAACTGACCTTCTGAATCATATTTCATTCCAAATCCA  4556
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71761  CCTGTTGGAGTCTATCCTAACTGACCTTCTGAATCATATTTCATTCCAAATCCA  71702

Qy  4557   CAAAACCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATTATTTTGTGTGTAG  4616
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71701  CAAAACCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATTATTTTGTGTGTAG  71642

Qy  4617   ACTTTCCTGATATTACACTGATTTGGAATATGAACAATTTATGGTTTCCTTTCGAA  4676
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71641  ACTTTCCTGATATTACACTGATTTGGAATATGAACAATTTATGGTTTCCTTTCGAA  71582

Qy  4677   GTAGTCAAGTCAAAGCAAAACAGCAAAAACAGCAAAAACTGTAAGACATAAAGAAATAGAGT  4736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71581  GTAGTCAAGTCAAAGCAAAACAGCAAAAACAGCAAAAACTGTAAGACATAAAGAAATAGAGT  71522

Qy  4737   GGAGCCGACTGAGAGATTAAAAATAAATATTTTATTAACAGGCAATTTGAAATA  4796
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71521  GGAGCCGACTGAGAGATTAAAAATAAATATTTTATTAACAGGCAATTTGAAATA  71462

Qy  4797   ATTTGTGCACTTCAGAATATTCTACAATAATATATATTATTTCCAATTTAATATCTTTAAG  4856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71461  ATTTGTGCACTTCAGAATATTCTACAATAATATATATTATTTCCAATTTAATATCTTTAAG  71402

Qy  4857   AAAATTACTATATTATATGTAAGTACATGTGCATGTGTTGAGGTAGGATATTTAACTCA  4916
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71401  AAAATTACTATATTATATGTAAGTACATGTGCATGTGTTGAGGTAGGATATTTAACTCA  71342

Qy  4917   ATAAAGTTATTTTCTTTATTCGGGTCAGGTCAAAGCTTCTAAGGGGATGTGAAAGGAT  4976
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71341  ATAAAGTTATTTTCTTTATTCGGGTCAGGTCAAAGCTTCTAAGGGGATGTGAAAGGAT  71282

Qy  4977   ATCTCTTTCTCTCTAGCTGAGAGAGTGAGTTCTAAGTTAAATATAATCAAGGAATTT  5036
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71281  ATCTCTTTCTCTCTAGCTGAGAGAGTGAGTTCTAAGTTAAATATAATCAAGGAATTT  71222

Qy  5037   CCCTGTCTTTGCTATTTGAGATTGTGACCACCAGGCGGTTGGCTGAAAGGGAAACTGA  5096
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71221  CCCTGTCTTTGCTATTTGAGATTGTGACCACCAGGCGGTTGGCTGAAAGGGAAACTGA  71162

Qy  5097   AGGGGCGGGGAGGGAGGGAGGGAAATAGATGAAAAAACAAAAACAAAACTTCCCTAAGCAG  5156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71161  AGGGGCGGGGAGGGAGGGAGGGAAATAGATGAAAAAACAAAAACAAAACTTCCCTAAGCAG  71102

Qy  5157   CTCTACAAAACATTTAGCCCCAGAAATAGTCACAGAAATCTCAAATCAACCAGTATC  5216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71101  CTCTACAAAACATTTAGCCCCAGAAATAGTCACAGAAATCTCAAATCAACCAGTATC  71042

Qy  5217   CAGATACAAGGAAGTGTTATGTGAGCTGGACACTCAGTTCAGT  5276
           |||||||||||||||||||||||||||||||||||||||||||
Db  71041  CAGATACAAGGAAGTGTTATGTGAGCTGGACACTCAGTTCAGT  70982

Qy  5277   TACAAAAGTCCAGGCTGCTGAAATAAACTCTGAAATTAAACTCTGATGCCAATCAC  5336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  70981  TACAAAAGTCCAGGCTGCTGAAATAAACTCTGAAATTAAACTCTGATGCCAATCAC  70922
```

```
Db  73140  TCTTTTCATAATTACTTATTTAAAATTGTTGTGCCATTTAATTTCTGTTCCTCTTAGCTT  73081

Qy  3177   AGTAACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTAAATGATAT  3236
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  73080  AGTAACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTAAATGATAT  73021

Qy  3237   TATTTAAAATACGTTAGGCTATAAACCTTTTAAATTTTTAAAAAATAGATGAGTGTGGT  3296
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  73020  TATTTAAAATACGTTAGGCTATAAACCTTTTAAATTTTTAAAAAATAGATGAGTGTGGT  72961

Qy  3297   GGCTCATGCTGTATCCCAACACTTTGGGAAGCGCGGGTCGGGGAGGATAGCTTGAGTCCA  3356
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72960  GGCTCATGCCTGTAATCCCAACACTTTGGGACGATAGCTTGAGTCCA  72901

Qy  3357   GCAGTTTGAGACCAGCCTCAGGGCAACACAGCCAAGACCCCATATCTAAAAAACAAACAAA  3416
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72900  GCAGTTTGAGACCAGCCTCAGGGCAACACAGCCAAGACCCCATATCTAAAAAACAAACAAA  72841

Qy  3417   ACAAAATTACCTGGTATGGTTGTGTCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGG  3476
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72840  ACAAAATTACCTGGTATGGTTGTGTCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGG  72781

Qy  3477   CAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTAC  3536
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72780  CAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTAC  72721

Qy  3537   ACTCAGTCTCGGGTGACAGTCGAAGAGCTGTCTCAAAATAATAAAATAAATAAAATAAC  3596
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72720  ACTCAGTCTCGGGTGACAGTCGAAGAGCTGTCTCAAAATAATAAAATAAATAAAATAAC  72661

Qy  3597   TTTTAAAAACAAAAATTAATTAAATTTAAAAACACAACACTAGATGTTTGCAAA  3656
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72660  TTTTAAAAACAAAAATTAATTAAATTTAAAAACACAACACTAGATGTTTGCAAA  72601

Qy  3657   TTGATTATTTGGGAGTCTATATCCCTGGAAGTTAATTTAAATATTTAAAGAGAGAGTTCTTC  3716
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72600  TTGATTATTTGGGAGTCTATATCCCTGGAAGTTAATTTAAATATTTAAAGAGAGAGTTCTTC  72541

Qy  3717   CTCATTTCCTAGAGCGTCGAATTGTAAATATCAGAGCTAGAGGAACACTAGGGCTCGC  3776
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72540  CTCATTTCCTAGAGCGTCGAATTGTAAATATCAGAGCTAGAGGAACACTAGGGCTCGC  72481

Qy  3777   CACTCCAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAAT  3836
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72480  CACTCCAAAGTGTGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAAT  72421

Qy  3837   GCAGAGTCTTAGGGCTCACGGGCACATTAAAACTTGAGAAGTCTGCACTAGAAATCTTCT  3896
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72420  GCAGAGTCTTAGGGCTCACGGGCACATTAAAACTTGAGAAGTCTGCACTAGAAATCTTCT  72361

Qy  3897   AGGTGCCTCACGGGCACATTAAAATCTAAAACTTCTGCACTCCACC  3956
           ||||||||||||||||||||||||||||||||||||||||||||||
Db  72360  AGGTGCCTCACGGGCACATTAAAATCTAAAACTTCTGCACTCCACC  72301

Qy  3957   TTTCATTATAAATGAATCACTTGGCTGTGTTGGTCACAGGAAATTGATTATTTTTAATTTC  4016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72300  TTTCATTATAAATGAATCACTT-GGCTGTGTGTGGTCACAGGAAATTGATTATTTTTAATTTC  72242

Qy  4017   AGAACCTTCTATTTAGGTCATCTATAATTCATCTATAATTGCTAATAGCACGGGAGAAAGCCAAACTCTTT  4076
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72241  AGAACCTTCTATTTAGGTCATCTATAATTGCTAATAGCACGGGAAGCCAAACTCTTT  72182

Qy  4077   AACTGCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTTACAT  4136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72181  AACTGCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTTACAT  72122

Qy  4137   TTTGTGGAGCAAGCTGTTTGATTTGGCTCGGGGCTCAGGCTGCCCTGTTTGTGAATTTCAC  4196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72121  TTTGTGGAGCAAGCTGTTTGATTTGGCTCGGGGCTCAGGCTGCCCTGTTTGTGAATTTCAC  72062

Qy  4197   AATTCACAGATGTTAGCCGCTTCTCGGGCTAAGTAAGGAGAGAATGTCAAGTTTTAAAT  4256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  72061  AATTCACAGATGTTAGCCGCTTCTCGGGCTAAGTAAGGAGAGAATGTCAAGTTTTAAAT  70981
```

```
Qy    5337  GACAGAGATCAGAGAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCT 5396
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70921  GACAGAGATCAGAGAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCT 70862

Qy    5397  ACTATACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATG 5456
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70861  ACTATACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATG 70802

Qy    5457  CCCAGCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCT 5516
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70801  CCCAGCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCT 70742

Qy    5517  AAAATCAAGAAACGCTCCGCCTCTTTGCAAATATGTATGAAGGAGAAGTGCCTAAACT 5576
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70741  AAAATCAAGAAACGCTCCGCCTCTTTGCAAATATGTATGAAGGAGAAGTGCCTAAACT 70682

Qy    5577  TCTATGTCTGATAGCATTTGTGTATATTTTATATAATTGTTCTCCGGCTTTATATCTATAT 5636
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70681  TCTATGTCTGATAGCATTTGACCCTATTGCTTTAGCCTCCCGGCTTTATATCTATAT 70622

Qy    5637  ACACAGGTATTGTGTATATTTTATATAATTGTTCTCCGT 5676
             |||||||||||||||||||||||||||||||||||||||
Db   70621  ACACAGGTATATGTGTATATTTTATATAATTGTTCTCCGT 70582
```

```
RESULT 7
AC135489/c
LOCUS         AC135489           160574 bp    DNA     linear     HTG 20-NOV-2002
DEFINITION    Rattus norvegicus clone CH230-436C20, WORKING DRAFT SEQUENCE, 4
              unordered pieces.
ACCESSION     AC135489
VERSION       AC135489.2  GI:25139116
KEYWORDS      HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE        Rattus norvegicus (Norway rat)
  ORGANISM    Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE     1  (bases 1 to 160574)
  AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
              Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
              Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
              Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
              Biswalo,K., Blair,J., Blankenburg,K., Blyth,C., Brown,M.,
              Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
              Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
              Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,A., Chu,J.,
              Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
              Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
              Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
              Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
              Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
              Fernandez,S., Finley,M., Flagg,N., Forbes,L., Forster,M., Foster,P.,
              Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
              Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guevara,W., Guevara,W.,
              Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
              Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
              Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
              Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
              Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
              Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
              Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
              Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
              Lorensuhewa,L., Louleeged,H., Lozado,R.J., Lu,X., Ma,J.,
              Maheshwari,M., Mahindarta,M., Mahmoud,H., Martin,R., Martinez,E.,
              Mangum,B., Mapua,P., Martin,E., Martin,R., Martinez,B.,
              Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
              Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
              Morgan,M., Morris,R., Morris,K., Munidasa,M., Murphy,M., Nair,L.,
              Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
              Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
              Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
```

```
              Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
              Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
              Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
              Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
              Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
              Shetty,A., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
              Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
              Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
              Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
              Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
              Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
              Williams,G., Willson,R., Wilczyk,R., Wooden,H., Worley,K.,
              Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
              Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
              Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
              Weinstock,G. and Gibbs,R.A.
  TITLE       Direct Submission
  JOURNAL     Unpublished
REFERENCE     2  (bases 1 to 160574)
  AUTHORS     Rat Genome Sequencing Consortium.
  TITLE       Direct Submission
  JOURNAL     Submitted (17-OCT-2002) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
REFERENCE     3  (bases 1 to 160574)
  AUTHORS     Rat Genome Sequencing Consortium.
  TITLE       Direct Submission
  JOURNAL     Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT       On Nov 20, 2002 this sequence version replaced gi:24080629.
              The sequence in this assembly is a combination of BAC based reads
              and whole genome shotgun sequencing reads assembled using Atlas
              (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
              in the feature table below represents a scaffold in the Atlas
              assembly (a 'contig-scaffold'). Within each contig-scaffold,
              individual sequence contigs are ordered and oriented, and separated
              by sized gaps filled with Ns to the estimated size. The sequence
              may extend beyond the ends of the clone and there may be sequence
              contigs within a contig-scaffold that consist entirely of whole
              genome shotgun sequence reads. Both end sequences and whole genome
              shotgun sequence only contigs will be indicated in the feature
              table.
              ----------  Genome Center
                 Center: Baylor College of Medicine
                 Center code: BCM
                 Web site: http://www.hgsc.bcm.tmc.edu/
                 Contact: hgsc-help@bcm.tmc.edu
              ----------  Project Information
                 Center project name: KBUG
                 Center clone name: CH230-436C20
              ----------  Summary Statistics
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 151333 bases at least Q40
                 Consensus quality: 153344 bases at least Q30
                 Consensus quality: 154259 bases at least Q20
                 Estimated insert size: 155447; sum-of-contigs estimation
                 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
              * NOTE: Estimated insert size may differ from sequence length
              *    (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
              * NOTE: This is a 'working draft' sequence. It currently
              *    consists of 4 contigs. The true order of the pieces
              *    is not known and their order in this sequence record is
              *    arbitrary. Gaps between the contigs are represented as
              *    runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              *    as soon as it is available and the accession number will
              *    be preserved.
              *      1    93077: contig of 93077 bp in length
              *  93078    93177: gap of unknown length
              *  93178   113865: contig of 20688 bp in length
              * 113866   113965: gap of unknown length
```

```
*       113966   158782: contig of 44817 bp in length
*       158783   158882: gap of unknown length
*       158883   160574: contig of 1692 bp in length.
FEATURES         Location/Qualifiers
   source        1..160574
                 /organism="Rattus norvegicus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10116"
                 /clone="CH230-436C20"
   misc_feature  complement(5637..6655)
                 /note="clone_boundary
                 clone_end:T7
                 site:
                 end_sequence:BZ203757"
   misc_feature  complement(90372..91276)
                 /note="clone_boundary
                 clone_end:Sp6
                 site:
                 end_sequence:BZ203758"
   misc_feature  93178..96221
                 /note="wgs_end extension
                 clone_end:Sp6"

ORIGIN

Query Match    11.8%;   Score 669.4;   DB 2;   Length 160574;
Best Local Similarity  68.5%;   Pred. No. 9.1e-118;
Matches 1192;  Conservative   0;  Mismatches 441;   Indels 108;   Gaps   15;

Qy   3981  GGCTGTGGTCACAGGAAATGATATTTTTAATTTCAGAACCTTCTATTAGGTCATCTA  4040
           |||  |||||| ||||||||  ||||| |||| ||||| ||||||||||||||||||
Db  65379  GGCCATAGCTACAGGACATTAGTTATTCTCAATTTAAGAACTTTGCAATCCGGTCATCTG  65320

Qy   4041  TATTTGCTAATAGCAGGGAAGAGCCAAACTCTTTAACTGCAARTTAA--CAAATCTATAA  4099
           ||||||||        |||||| ||||||||||||| ||||| ||||   |||||||| ||
Db  65319  TATTTGCT----CAGAGAGAAAGTCAAACACTTCAAATTCCAATGAAGCAAATCTATAA  65265

Qy   4100  TTAATTAGTTAAGCAATCTTCCCTTGAAGTTTGAGAGCCAAGCTGTTTGATT  4159
           ||||||||||| ||| || ||  ||||||||||||||| || ||||||| ||||
Db  65264  TTAATTAGTGAAATGATCTTTGCTTCAAGCTTCACATTTTTAGAAAAAGCTCTCTGATT  65205

Qy   4160  TGGCTGGGGCTCAGGCCGGCCTGTTTGTGAATTCACAATTCACAGATGTTAGCGCGCTCT  4219
           |||||||||||||||||| |||||| || ||| || |||||||||| |||||||| ||
Db  65204  TGGCTGGGGCTCAGGCCCCCATGTGTGTGAGTTTCACAATGTTAGCACACGCT  65145

Qy   4220  CGGGCTAAGTAAAGAGAGAAATGTCAAGTTTAAATAGCTTCTCCCTTCCATCCTGGCT  4279
           || |||||| ||| ||||||||||||||||||||| |||| ||||| |||||||||
Db  65144  CAGTCGAGGAAGGAAGAATGTCAAGTTTAAATAGCTTCTC--------CCTGGCT  65093

Qy   4280  GAAGCAACAAATAAAATATTTTATGAAACACATTTGAGTTAGATTACTTACAGGGAA  4339
           | ||||||||||||||||||| || |||||| ||||||| |||||||| |||||||||
Db  65092  AGAACAACAAATAAAAACATCTTTATG-AACACATTTCCTTACAGGAAATTCTTACAAGAAA  65034

Qy   4340  ATGTCAAATTTCTCTGAAAGGGCTTTGAGATTGTCTCACAATCTACTACTGATGTC  4399
           ||| ||||||||           ||||||||||||||||||| |||||| ||||||||
Db  65033  ATGTCAGATTTCT--------CTTCAGCTTGTCTCACAACTTTGACATTTTCTGATGTC  64983

Qy   4400  ACCTATTTACAGGTGTGTCCTGTGCACTAGGGCCTAGTGGGAGTGAAGGGAAGATGTGAACTCACCATGT  4459
           || ||| |||||||||||  |||  ||||||   ||| |||||| ||| || |||| ||||||||||||
Db  64982  ACTAGTTTGCAGGTGT---CTATGACTACCGCATGAATGG-AGATATGTACTTAGCATGT  64927

Qy   4460  TAGTGACCGTTAGATACACAGAGTGTTTTTTTCCCCCCGTTGGGAGTCTATCCTTAACTG  4519
           ||||||||| |||||| ||| ||||       ||||| ||| ||| |||||| |||||
Db  64926  CAGCGACATTTGGGCACACGG------CTCTATTTCCCTCTTGGAGTCTATTCATTCTAAATT  64874

Qy   4520  AGCTTCTGAATCATATTTCATTCAATTCATTCAATTTCTAAATTCTAAATCCACAGACC  4579
           ||||||||||| ||||| || ||||| || ||||| |||||| |||| ||||| |||||
Db  64873  AGCTTCAGAATCATATTTCATTCAATTCATTCTAAATCCACAGGATTTGTTTACGGC  64814

Qy   4580  CCATATTCAGAAAGGAAAATAAATTATTTTGTGTAGACTTT-CCTGATATTACACTGAT  4638
           |||||| |||||| |||| ||||| |||||| || || |||  |||||||| ||| ||||
Db  64813  CTACATTCAGAAAGAAACAAATTCATCATCATGGAGACTTTGCCTGATAATATATTTAAT  64754

Qy   4639  TTGGGAATATATGAACAATTTTATGGTTTCCTTTCGAAGTAGGTCAAGTCAAGTCAAAGCAAAAC  4698
```

```
Db  64753  TCTTCACTATATAAACAAATCGTAAGACATTTTTAGAGTAGGTAAGTCAAGGGACTGAGAGATTAAAA  64694
Qy   4699  CAAAAACAGCAAAAACTGTAAGACTAAAGAATAGAGTGAGGCCGACTGAGAGATTAAAA  4758
           |||| |  |||  ||||| ||||| ||||| |||| ||| ||||||||||| ||||||
Db  64693  CAACTAAAGTGAGACGTAAGAACATAAGACACTGATAGAGTTTGAGGTCAACTTTAT  64634
Qy   4759  TAAACTAGAATATTTTTATTAACAGGCAATTGAAAATAATTGTGCACTTCAGAATATTC  4818
           |||| |||| ||| |||||| || ||||||||||| ||||  ||| |||||||||| ||
Db  64633  TTTTCCCATTGACGGGCTATCTGAAATAATTCTAAATAATAAAAACAAT  64574
Qy   4819  TACAATAATATATTCCAATTTAATATCTTTAAGAAAATTACTATATATATGTAA  4878
           |||| || ||||| ||||||  || |||| ||| ||||||  ||||| ||||| ||
Db  64573  TAAGAGCATTATGAAAGAAAAACATTAACGTATTAACTCCTATTTTCTTTTATT  64514
Qy   4879  GTACATGTGCATGTGTTGAGTAGGATATTTAACTCAATAAAGGTTATTTCTTTTATT  4938
           || || |  |||  || |||||| ||||||| ||||||| ||||||  ||||| |||||
Db  64513  ATACTATAACTTACATCTGTATGGATAAAAATCACCAC--AGGCTATTTTTTTAATT  64457
Qy   4939  CGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCCTTTCTCT----------  4988
           || ||||| ||||||||||          ||| | ||| ||| |||||||||
Db  64456  TGGAGAAGAAAATATTTCTAAATACATAGGCCATTCTCTCTCTCTCTCTCTCTCTCTCTCTCT  64397
Qy   4989  ---------------------------------------TAGCTGAGGAGAGTGAGTTCTAAG  5015
                                                  |||||||||||||||||||||
Db  64396  CTCTCTCTCTCTCTCTCTCTCCCGTGTCTTGCTATTTGACGAGTTGACCACAACAGCCG  64337
Qy   5016  TTAAATATAATCAAGGAATTTCCCGTCTTGCTATTTGAGATTGTGACCAGGTGA  5075
           |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  64336  TTAAATATAATCAAGGAATTTCCCGTCTTGCTATTTGAGATTGTGACCACAGCGCG  64277
Qy   5076  GTTGGGCTGAAAGGGAAACTGAAGGGCGGGAGGGGAGGGAAATAGAATGAAAAAACAAAACA  5135
           |||||||||||||||||||||||||||||| |||||
Db  64276  GTTGGGCTGAAAGGGAGGCTGAAGGGGTTGGGGAG--------GGAAGAGAGGGGG  64226
Qy   5136  AAACAAAACTTCCCTAAGCAGCTCTACAAAAACATTTTAGCCCCAGAAAGTCACAGAA  5195
           ||| ||||||| ||| |||||  || |||||   ||| ||| |||  |||| ||||||
Db  64225  AAAAAATACTCTCCTAAGCGTCTCTGCCAAACAGTCTGCCACAGTCTGACAGGAG  64166
Qy   5196  TCCTCAAATCAAACCAGTATCCAGATACAAGGAAGTGTATGTAGCTGGAGCAGGGTGA  5255
           |||||||||||||||| ||||| ||||||||||||| ||||| ||||| || |||||
Db  64165  CTCTCAAATCAAATCAGATACAGATACAAGGAGATGTTATTCAGTTGGAACAGGGGGA  64106
Qy   5256  CACTCATCAGCTCAGTTCAGTTACAAAAGTCAGGCTGCTGAAATTAAACTCTGATGCCA  5315
           ||  ||         | |||||  ||| |||  || ||||||||| || ||||||||
Db  64105  CATTTATTAACTCAGT------GACAAGTCGTCTGGCTCCCGTGATTAAACTCTGATGCCA  64053
Qy   5316  TTCATGCCAGCATCCAATCACGACAGAGTCAGAGATGCTCAGCTCCAAA----------  5375
           |||||||||||| ||||||||| ||||||| |||||  ||||||||||
Db  64052  TTCATGCCAGCACCCAATCACACGGACAAGACCAGAAGTCAGATGCCTACAAAT------  63998
Qy   5376  TTGCCAACAACAAGTGTGGCTACTATGCGTCAAGGACTCTGAAGCCGTGAGACGAGGGGGA  5435
           |||||||||||||||||||||| ||    |||  ||| |||| |||||| ||||||
Db  63997  ----TGCCAACAAGTGTGGCCCACTCTCTATGTGTCAAGGGCTCTAAAACTGGGCAGA-GAGGA  63943
Qy   5436  AGAACAACAGTAGAGAGGATGCCCAGCTGGTAAGAATCGAGTGTTTTAGTC  5495
           || ||||| |||| ||||||||||||| |||  |||| ||||||  ||||||
Db  63942  AGAACAGCTTTACAGAGGATGCCCAGCTGGTAAGAGCTGACTGTTTATGATGCTCCGGTC  63883
Qy   5496  AATTGATGAATCTCATTGGCTAAATGCTGGAGTCTATCCTTAACTG  5555
           |  ||||||||| |||||||| ||||| |||||||| |||||||||
Db  63882  ACTTGAAGACCGTGATTGGCTGAAAGGAAGAAAATGTCTCCGCCTCTTAGCAAATCTGTGTA  63823
Qy   5556  AAGGAGGAAGTGCCTAAACTTCTATGTCTGATAGCATTTGACCCTATTTGCTTTTAGCCT  5615
           || ||| |||||| ||||||||||| |||||  | ||||||||||| |||| |||| ||||
Db  63822  AAGGGGGGAAGTGTCTAAACTTCTAAACTTGTCTATGCATGTGATTGACCCTATTTGCTTTCAGCCT  63763
Qy   5616  CCCGGCTTTATATCTATATATACACAGGTATTTGTGTATATTTTATATAATTGTTCCG  5675
           || || |||||| ||| |||||||||||  |||| ||||| ||||| |||| ||||
Db  63762  CCTGGCTATATACCTAGATATTCTCAGGTGTATATGTGTATATTTTATGAATTGTCCCCA  63703
Qy   5676  T  5676
           |
Db  5676
```

```
Db        63702 T 63702

RESULT 8
AC113784/c
LOCUS       AC113784              284005 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-98F2, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC113784
VERSION     AC113784.5  GI:25072720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 284005)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,D., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,K., Ceasar,H., Center,A., Chu,J.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemen,O., Okwuonu,O., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,V., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk.R., Wooden,H., Worley,K.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 284005)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
REFERENCE   3  (bases 1 to 284005)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23195097.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------------------- Genome Center
                 Center: Baylor College of Medicine
                 Center code: BCM
                 Web site: http://www.hgsc.bcm.tmc.edu/
                 Contact: hgsc-help@bcm.tmc.edu
            ------------------- Project Information
                 Center project name: GSGN
                 Center clone name: CH230-98F2
            ------------------- Summary Statistics
                 Assembly program: PhrAp; version 0.990329
                 Consensus quality: 249691 bases at least Q40
                 Consensus quality: 252542 bases at least Q30
                 Consensus quality: 254124 bases at least Q20
                 Estimated insert size: 252183; sum-of-contigs estimation
                 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            -------------------
            * NOTE: Estimated insert size may differ from sequence length
            *      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            *      consists of 3 contigs. The true order of the pieces
            *      is not known and their order in this sequence record is
            *      arbitrary. Gaps between the contigs are represented as
            *      runs of N, but the exact sizes of the gaps are unknown.
            *      This record will be updated with the finished sequence
            *      as soon as it is available and the accession number will
            *      be preserved.
            *         1    9564: contig of 9564 bp in length
            *         9565       9664: gap of unknown length
            *         9665     273939: contig of 264275 bp in length
            *       273940     274039: gap of unknown length
            *       274040     284005: contig of 9966 bp in length
FEATURES             Location/Qualifiers
     source          1..284005
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-98F2"
     misc_feature    1..1035
                     /note="wgs end extension
                     clone end:T7"
     misc_feature    7353..8173
                     /note="clone boundary
                     clone end:T7
                     site:"
                     end sequence:BH306729"
     misc_feature    9665..11350
                     /note="wgs contig"
     misc_feature    14183..15278
                     /note="wgs contig"
     misc_feature    15329..16533
                     /note="wgs contig"
     misc_feature    complement(263454..264092)
```

```
        /note="clone boundary
        clone_end:Sp6
        site:
misc_feature   end_sequence:BH306731"
        274040..275276
        /note="wgs_end_extension
        clone_end:Sp6"

ORIGIN

Query Match    11.8%;    DB 2;    Length 284005;
Best Local Similarity   68.5%;   Pred. No. 8.6e-118;
Matches 1192;   Conservative   0;   Mismatches 441;   Indels 108;   Gaps 15;
```

```
Qy    3981  GGCTGTGGTCACAGAAATGATTATTTTTAATTTCAGAACCTTCTATTTAGGTCATCTA  4040
            ||||  ||||||||||  |||||||    |||  |||||| ||  |  |||||||||||
Db   72189  GGCCATAGCTACAGGACATTAGTTATTCTCAATTTAAGAACTTTCGAATCGGGTCATCTG  72130

Qy    4041  TATTTGCTAATAGCAGGGAAGAAAGCCAAACTCTTTAACTGCAATTAA-CAAATCTATAA  4099
            |||||||||||      |||  ||||  ||  ||||||||||  ||  ||||||||||||
Db   72129  TATTTGCT-----CAGAGAAGAAAGTCAAACACTTCAATTCCAATGAAGCAAATCTATAA  72075

Qy    4100  TTAATTAGTTAAGCAATCTTCCCTTAAGTTTTACATTTGTGGAGCAAGCTGTTTGATT  4159
            ||||||||  || ||||||||  |||| ||||||| |||  |||  |||| ||  ||||
Db   72074  TTAATTAGTGAAATGATCTTTGCTTCAAGCTTCACATTTTTAGAAAAAGCTCTCTGAATT  72015

Qy    4160  TGGGTGGGGCTCAGGCCGGCCTGTTTGTGATTTCACAATTCACAGATGTTAGCCGCTCT  4219
            |||||||||||||||||||     |||||| |||||||||||||| || |||||||||
Db   72014  TGGGTGGGGCTCAGGCCCCCATGTGTGTGAGTTTCACAGATGTTCACAGATGTTAGACACGCT  71955

Qy    4220  CGGGCTAAGTAAAGGAAGAATGTCAAGTTTTAAATAGTTCTCCCTTCCATCCTGGCT  4279
            |||| |||||||||||||  ||||| ||||||||||||  |  |      |||||||
Db   71954  CAGTCGAGGAAAGGAAGAATGTCAAGTTTTAAATAGCTTCTC-------CCTGGCT  71903

Qy    4280  GAAGCAACAAATAAAATATTTTTATGAAACACATTTGAGTTAGATTTACTTACGGGAA  4339
            ||||||||||||||||    ||||  ||| |||||||||| ||||| |||||||||||
Db   71902  AGAACAACAAATAAAACATCTTTATG-AACACATTTTTATG-AACCAGAATTTCTTACAAGAAA  71844

Qy    4340  ATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAACTTGTCACATCTACTGATGTC  4399
            |||||||||||||       |||||||| ||||| |||||||||||||| ||| ||||||
Db   71843  ATGTCAGATTTCT-------CTTCAGCTTGTCTCACAACTTTGACATTTCTGACATGTC  71793

Qy    4400  ACCTATTTACAGGGTGTGTCCTGTCGACTAGGGGGTGAAGGAAGATGTGAACTCACCATGT  4459
            |||||||  |||  || || |      ||||||||||  |||||||||||||  |||||||
Db   71792  ACTAGTTTGCAGGTGT---CTATGACTACGGCATGAATCG-AGATATGTACTTAGCATGT  71737

Qy    4460  TAGTGACCGTTAGATACAGAGTGTTTTTTCCCCCTGTTGGAGTCTCATCCTAACTG  4519
            |||||||||||       ||  |||||| | |||| ||||||| ||| ||||| ||
Db   71736  CAGCCGACACATTTGGGCACACGG------CTCTATTTCCCTCTTGGGACTCCTATTCTAATT  71684

Qy    4520  AGCTTCTGAATCATATTTCATTCAAATTTCCAAATCCACAGAAACCAGGATAAGTTTACAGC  4579
            |||||||||| |||||||||| | |||||   ||| |||||||||||||| |   ||||||
Db   71683  AGCTTCTGAATCATATTTCATTCAAATTTCTAAATCCACAGAACCAGGATTGGTTTACGGC  71624

Qy    4580  CCATATTCAGAAAGGAAGAAATAAATTATTTTGTGTGTAGACTTT-CCTGATATTACTACTGAT  4638
            ||||||||||||||  |||||||||        |   ||||| ||| |||||| |||||||||
Db   71623  CTACATTCAGAAAGGAAAGAAACAAATTCATCATCAGAGACTTGCCTGATAATAATATTAAT  71564

Qy    4639  TTGGGAATATATGAACAAATTTATGTTCCTTTCGAAGTAGGTCAAGTCAAAGCAAAAC  4698
            |||   |||| |  ||||||||||| |  ||||||| ||| |||||||||| |||||
Db   71563  TCTTCACTATATAAACAAAATGCCAGATTTTTAGAGTAGTAAGTCAAGGGAGACTAAGAG  71504

Qy    4699  CAAAAACAGCAAAAACTGTAAGACATAAAGACATTACTAGAGTCGAGCCCGACTGAGAGATTAAAA  4758
            |||                |||     || | ||  | | |||
Db   71503  CAACTAAAGTGAGACGTAAGGAGTAAGACAGAGTAGAATTGGCACTTCGTGCACTTCAGATATTC  71444

Qy    4759  TAAACTAGAGATATTTTTTATTAACAGGCAATTGAAATAATTTGTGCACTTCAGATATTC  4818
            |||                |||||||||||||||||||  ||||||||||  ||
Db   71443  TTTTCCCATTGACACGGGCTATCTGAAATAATTCTAAATAATATAAATCCAAATAATAAAAACAAT  71384

Qy    4819  TACAATAATATATTATTTCCAATTTTAAATATCTTTAAGAAAAATTACTATTATATGTAA  4878
            |||       ||||     |    |||||||  |||||||||||| ||  |||  |||||
Db   71383  TAAGAGAGCATTATGAAGAAGAAACATTAACGTATTAACTCCTATTTTAATATTCACATTAAA  71324
```

```
Qy    4879  GTACATGTGCATGTGTTTGAGGTAGGAGATATTTAACTCAATAAAGTTATTTTCTTTTATT  4938
            ||||            |   |   |   |  || |||||||||  ||||   |  |||||||
Db   71323  ATACTATAACTTACATCTGTATGGATGAAAATCACCAC---AGGCTATTTTTTTTTAATT  71267

Qy    4939  CGGGTCAGGCAAAGCTTCTAAGGATGTGAAAGGGATATCTCTTTCTCT---------  4988
            ||||||||||||| ||||||||||||||||| |||||||||||||||||
Db   71266  TGGAGAAGAAAATATTTCTAAATACATAGGCATTCTCTCTCTCTCTCTCTCTCTCTCT  71207

Qy    4989  --------------------------------TAGTCTGAGGAGGAGTCAGTTCTTAAG  5015
                                            |||||||||||||||||||||| ||||
Db   71206  CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGCTCTGTCGAAACCAAGAGTCAGTTCTAGG  71147

Qy    5016  TTAAATATAATCAAGGAATTTCCCTGTCTTGCTATTGACATTGTGACCACAACAGGCCG  5075
            ||||||||||||||||||||||||||||  |||||||||||||||||| |  |||||||
Db   71146  TTAAATATAATCAAGGAATTTCCCTGTCTTGCTATTGACATTGTGACCACAACAGGCG  71087

Qy    5076  GTTGGCTGAAAGGGAAACTGAAGGGCGGGGAGGCGGGAGGGAAATAGATGAAAAAACAAAACA  5135
            |||||||||||||||||||||||     ||||||| |||||||
Db   71086  GTTGGCTGAAAGGGAAGGGAGGCTGAAGGGGTTGGGAGG--------GGAAGAGAGGGGG  71036

Qy    5136  AAACAAAACTTCCCTAAGCAGCTTCACAAACATTTTAGCCCCAGAAATAGTCACAGAAA  5195
            |||||||||||||||||||  |||| ||  || ||  ||  |||||||||||||||||||
Db   71035  AAAAAATACTTCCCTAAGCAGTCTCTGCCAAACAGTCTAGCCAGCAGCAGCACGAGAG  70976

Qy    5196  TCCTCAAATCAAACCAGTATCCAGATACAAGGAAGTGTTATGTAGCTGGAGCAGGTGGA  5255
            ||||||||| ||||| |||| ||||||||||||||||| |||||||||||||  ||||||
Db   70975  CTCTCAAATCAAATCAGAATACAGATACAGGAGAGATACAAGGAGAGTTGGAACAAGGGGA  70916

Qy    5256  CACTCATCAGTCAGTTCAGTTACAAAAGTCCAGGTGCTGAAAATTAAACTCTGATGCCA  5315
            ||        || ||||     |      |    |||||  | || ||||||||||||||
Db   70915  CATTTATTAACTCAGT-------GACAAGTCCTGGCTCGTGTGATTAAACTCTGATGCCA  70863

Qy    5316  TTCATGCCAGCATCCAATCACGACAGAGATCAGAGAGGCAAGAGTCAGAGATGCCTCCAAA  5375
            ||||||||||||||||  ||||| ||  ||| || ||||| ||||| ||||||||| ||
Db   70862  TTCATGCCAGCACCCAATCACGGACAAGAGAAGAGAAGTCAGAGATGCCTACAAAT------  70808

Qy    5376  TTGCCAACAACAAGTGTGGCTACTATACGTCAAGGAACACTGAGCCGTGAGAGAGGGGGA  5435
                   |||||||| ||||||||||   ||||| ||  |||| | |||||| |  |||||
Db   70807  ----TGCCAACAAGTGTGGCCACTGTCATGTCAAGGGCTCAAACTGGCAGA-GAGGA  70753

Qy    5436  AGAACAACAGTAGAGAGGATGCCCAGCTGGTAAGAATCGAGTGTTTAGTTTTAGTC  5495
            |||||| |||   || ||||||||||||||  ||| ||| | ||||| ||||| |||
Db   70752  AGAACAGCTTTACAGGAGGATGCCCAGCTGGTTGTCAGTGTTTATGATGCTCCGGTC  70693

Qy    5496  AATTGATGAATCTCATTGGCTAAAATCAAGAAACAGCTCCGCCCTCTTGCAAATATGTATG  5555
            | || ||  || ||  |||||| |  ||||||| || | ||||||||||  ||||| |||
Db   70692  ACTTGAAGACGCTGATTGGCTGAAAGGAAGAAAATGCTCCGCCCTCTTAGCAAATCTGTGTA  70633

Qy    5556  AAGGAGAAGTGCTCAAACTTCTATGTCTGATAGCATTTGACCCTATTGCTTTTAGCCT  5615
            |||   |||||| |||||||||||||| ||| ||||| |||||| ||| ||||||||||
Db   70632  AAGGGGGAAGTGTCTCAAACTTCTATGTCTGATGGCATTTGACCCCTATTGCTTCAGCCT  70573

Qy    5616  CCCGGGTTTATATCTATATACACAGGTATTTGTGTATATTTATATAATTGTTCTCCG  5675
            |||||| ||| |  ||||  |||||| ||||  || |||||| |||||||| ||||||
Db   70572  CCTGGGCTATATACCTAGATATTTCTCAGGTGTATATGTATATTTTATAGAATTGTTCCCCA  70513

Qy    5676  T  5676
            |
Db   70512  T  70512
```

REFERENCE   Rattus.
             1 (bases 1 to 204313)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
             Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
             Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
             Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
             Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
             Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
             Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
             Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
             Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
             Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
             Delgado,O., Denson,S., Deramo,C., Ding,J., Dinh,H., Divya,K.,
             Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
             Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
             Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
             Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
             Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
             Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
             Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
             Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
             Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
             Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
             Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
             Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
             Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
             Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
             Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
             Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
             Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
             Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
             Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
             Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
             Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
             Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
             Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
             Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
             Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
             Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
             Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
             Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
             Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
             Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
             Taylor,T., Thomas,N., Thomas,J., Tingey,A., Trejos,Z., Usmani,K.,
             Valas,V., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
             Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
             Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
             Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,V.,
             Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
             Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
             Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 204313)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (05-JUL-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 204313)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 20, 2002 this sequence version replaced gi:23907953.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/).Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
--------------- Genome Center
   Center: Baylor College of Medicine
   Center code: BCM
   Web site: http://www.hgsc.bcm.tmc.edu/
   Contact: hgsc-help@bcm.tmc.edu
--------------- Project Information
   Center project name: G2GF
   Center clone name: CH230-244D17
--------------- Summary Statistics
   Assembly program: Phrap; version 0.990329
   Consensus quality: 190230 bases at least Q40
   Consensus quality: 193082 bases at least Q30
   Consensus quality: 194928 bases at least Q20
   Estimated insert size: 19768; sum-of-contigs estimation
   Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*   1   71914: contig of 71914 bp in length
*   71915   72014: gap of unknown length
*   72015   203206: contig of 131192 bp in length
*   203207   203306: gap of unknown length
*   203307   204313: contig of 1007 bp in length.
FEATURES   Location/Qualifiers
     source   1..204313
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-244D17"
     misc_feature   complement(4101..4830)
                     /note="clone_boundary
                     clone_end=T7
                     site:
                     end_sequence:BZ158614"
     misc_feature   complement(69055..69921)
                     /note="clone_boundary
                     clone_end=Sp6"
                     site:
                     end_sequence:BZ158615"
     misc_feature   72015..73525
                     /note="wgs_end_extension
                     clone_end=Sp6"
     misc_feature   191056..192396
                     /note="wgs_end_extension
                     clone_end=Sp6"
ORIGIN
       Query Match   11.7%;   Score 662; DB 2; Length 204313;
       Best Local Similarity 68.2%;   Pred. No. 2.3e-116;
       Matches 1187; Conservative   0; Mismatches 446; Indels 108; Gaps 15;

Qy   3981 GGCTGTGGTCACAGGAAAATTGATTATTTTTAATTTCAGAACCTTCATTAGGTCATCTA 4040
            |||||||||||||||||||||||||||||||||||||||||| |||| |||||||||||
Db   63951 GGCCATAGCTACAGGACATTAGTTATTATTCTCATTTAAGAACTTTGCAATCGGGTCATCTG 63892

Qy   4041 TATTTGCTAATAGCACGGAAGAAAGCCAAACTCTTTAACTGCAATTAA-CAAATCTATAA 4099
            ||||||| |||||||||| ||||||||||||||||||| ||||||| ||||||||||
Db   63891 TATTTGCT----CAGAGAAGAAAGTCAAACACTTCAATTCCAATGAGCAAAATCTATAA 63837

Qy   4100 TTAATTAGTTAAGCAATCTTCCCTTAAGTTTACATTTGTGGAGCAAGCTGTTTGATT 4159

RESULT 10
AC113020
LOCUS   AC113020   214853 bp   DNA   linear   ROD 04-OCT-2003
DEFINITION Mus musculus chromosome 16, clone RP23-186N8, complete sequence.
ACCESSION   AC113020
VERSION   AC113020.9   GI:37515087
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 214853)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Mus musculus chromosome 16, clone RP23-186N8
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 214853)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

```
                  Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
                  Topham,K., Travis,M., Travis,N., Trigilio,J., Vassiliev,H.,
                  Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
                  Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE             Direct Submission
JOURNAL           Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
                  Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE         3  (bases 1 to 214853)
AUTHORS           Birren,B., Nusbaum,C. and Lander,E.
TITLE             Direct Submission
JOURNAL           Submitted (05-SEP-2003) Whitehead Institute/MIT Center for Genome
                  Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE         4  (bases 1 to 214853)
AUTHORS           Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
                  Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
                  Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
                  Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
                  Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
                  Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
                  Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
                  Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
                  Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
                  Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
                  Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
                  Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
                  Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
                  O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
                  Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
                  Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
                  Spence,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
                  Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
                  Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
                  Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE             Direct Submission
JOURNAL           Submitted (04-OCT-2003) Whitehead Institute/MIT Center for Genome
                  Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT           On Oct 4, 2003 this sequence version replaced gi:34482079.
                  All repeats were identified using RepeatMasker:
                  Smit,A.F.A. & Green,P. (1996-1997)
                  http://ftp.genome.washington.edu/RM/RepeatMasker.html
                  ----------- Genome Center
                  Center: Whitehead Institute/ MIT Center for Genome Research
                  Center code: WIBR
                  Web site: http://www-seq.wi.mit.edu
                  Contact: sequence_submissions@genome.wi.mit.edu
                  ----------- Project Information
                  Center project name: L23503
                  Center clone name: 186_N_8
                  -----------
FEATURES             Location/Qualifiers
     source          1..214853
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /map="16"
                     /clone="RP23-186N8"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1..51
                     /note="wgs end_extension
                     clone_end:SP6"
     misc_feature    52..57
                     /note="clone_boundary
                     clone_end:SP6"
                     site:EcoRI"
     repeat_region   1042..1080
                     /rpt_family="(CATATA)n"
     repeat_region   1179..1367
                     /rpt_family="B3"
     repeat_region   2436..2578
                     /rpt_family="L1M4c"
     repeat_region   2651..2737
                     /rpt_family="L1M4c"
     repeat_region   complement(2850..3351)
                     /rpt_family="RMER19B"
     repeat_region   complement(3350..3516)
                     /rpt_family="B3"
     repeat_region   3527..3891
                     /rpt_family="L1M4c"
     repeat_region   4170..4563
                     /rpt_family="L1M4c"
     repeat_region   4873..4952
                     /rpt_family="B1_MM"
     repeat_region   4968..5018
                     /rpt_family="B1_MM"
     repeat_region   5056..5415
                     /rpt_family="(TGAA)n"
     repeat_region   5554..5810
                     /rpt_family="MTD"
     repeat_region   6353..6500
                     /rpt_family="L1M4c"
     repeat_region   complement(6731..7895)
                     /rpt_family="Lx8"
     repeat_region   7896..7919
                     /rpt_family="Lx7"
     repeat_region   complement(7920..8304)
                     /rpt_family="(TTTTG)n"
     repeat_region   complement(8309..9900)
                     /rpt_family="Lx7"
     repeat_region   complement(9899..10500)
                     /rpt_family="Lx4"
     repeat_region   10516..10639
                     /rpt_family="Lx4"
     repeat_region   10634..11419
                     /rpt_family="L1"
     repeat_region   11420..11615
                     /rpt_family="L1"
     repeat_region   11616..13898
                     /rpt_family="B2_Mm2"
     repeat_region   complement(13899..14142)
                     /rpt_family="L1"
     repeat_region   14447..17150
                     /rpt_family="LTRIS_MM"
     repeat_region   17251..17303
                     /rpt_family="Lx8"
     repeat_region   17523..17580
                     /rpt_family="(TGG)n"
     repeat_region   17590..17667
                     /rpt_family="AT_rich"
     repeat_region   17669..18434
                     /rpt_family="(GAAA)n"
     repeat_region   18558..18593
                     /rpt_family="Lx8"
     repeat_region   18855..19268
                     /rpt_family="AT_rich"
     repeat_region   21016..21156
                     /rpt_family="BGLII"
     repeat_region   21157..21199
                     /rpt_family="B1_MM"
     repeat_region   complement(21371..21767)
                     /rpt_family="(CAAA)n"
     repeat_region   complement(21768..21878)
                     /rpt_family="LLME"
     repeat_region   21869..22524
                     /rpt_family="L1"
     repeat_region   22533..22555
                     /rpt_family="L1_MM"
     repeat_region   complement(22556..25200)
                     /rpt_family="(A)n"
     repeat_region   complement(25514..25708)
                     /rpt_family="L1"
     repeat_region   26244..26293
                     /rpt_family="B3"
     repeat_region   26294..26369
                     /rpt_family="MT2B"
     repeat_region   26295..26369
                     /rpt_family="B1_MM"
```

repeat_region   /rpt_family="ID_B1"
26370..26532
repeat_region   /rpt_family="(GAAA)n"
26533..26553
repeat_region   /rpt_family="ID_B1"
26674..26921
unsure   26674..26921
/note="single clone coverage"
repeat_region   26678..26712
/rpt_family="(TTTG)n"
complement(26715..26836)
/rpt_family="B1_MM"
repeat_region   26898..26945
/rpt_family=" (TC)n"
repeat_region   26945..26989
/rpt_family=" (TG)n"
complement(26992..27264)
/rpt_family="B4A"
repeat_region   30432..30458
/rpt_family=" (T)n"
repeat_region   30639..30668
/rpt_family=" (CCCCT)n"
complement(31137..31532)

Query Match   11.4%;   Score 649;   DB 10;   Length 214853;
Best Local Similarity   68.4%;   Pred. No. 6.9e-114;
Matches 1210;   Conservative   0;   Mismatches   420;   Indels   139;   Gaps   17;

[DNA sequence alignment blocks — Qy (Query) and Db (Database) pairs with numbered positions]

```
RESULT 11
AC091240/c
LOCUS       AC091240      56268 bp    DNA     linear   HTG 07-APR-2001
DEFINITION  Mus musculus clone RP23-16N1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC091240
VERSION     AC091240.1  GI:13560417
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 56268)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-16N1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 56268)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,P., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code:WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
            ------------- Project Information
            Center project name: L13295
            Center clone name: 16_N_1
            -------------
          * NOTE: This record contains 70 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *       1     713: contig of 713 bp in length
          *     714     813: gap of 100 bp
          *     814    1517: contig of 704 bp in length
          *    1518    1617: gap of 100 bp
          *    1618    2314: contig of 697 bp in length
          *    2315    2414: gap of 100 bp
          *    2415    3091: contig of 677 bp in length
          *    3092    3191: gap of 100 bp
          *    3192    3877: contig of 686 bp in length
          *    3878    3977: gap of 100 bp
          *    3978    4658: contig of 681 bp in length
          *    4659    4758: gap of 100 bp
          *    4759    5468: contig of 710 bp in length
          *    5469    5568: gap of 100 bp
          *    5569    6273: contig of 705 bp in length
          *    6274    6373: gap of 100 bp
          *    6374    7094: contig of 721 bp in length
          *    7095    7194: gap of 100 bp
          *    7195    7890: contig of 696 bp in length
          *    7891    7990: gap of 100 bp
          *    7991    8716: contig of 726 bp in length
          *    8717    8816: gap of 100 bp
          *    8817    9527: contig of 711 bp in length
          *    9528    9627: gap of 100 bp
          *    9628   10338: contig of 711 bp in length
          *   10339   10438: gap of 100 bp
          *   10439   11128: contig of 690 bp in length
          *   11129   11228: gap of 100 bp
          *   11229   11949: contig of 721 bp in length
          *   11950   12049: gap of 100 bp
          *   12050   12758: contig of 709 bp in length
          *   12759   12858: gap of 100 bp
          *   12859   13572: contig of 714 bp in length
          *   13573   13672: gap of 100 bp
          *   13673   14357: contig of 685 bp in length
          *   14358   14457: gap of 100 bp
          *   14458   15167: contig of 710 bp in length
          *   15168   15267: gap of 100 bp
          *   15268   15949: contig of 682 bp in length
          *   15950   16049: gap of 100 bp
          *   16050   16738: contig of 689 bp in length
          *   16739   16838: gap of 100 bp
          *   16839   17545: contig of 707 bp in length
          *   17546   17645: gap of 100 bp
          *   17646   18372: contig of 727 bp in length
          *   18373   18472: gap of 100 bp
          *   18473   19206: contig of 734 bp in length
          *   19207   19306: gap of 100 bp
          *   19307   20017: contig of 711 bp in length
          *   20018   20117: gap of 100 bp
          *   20118   20835: contig of 718 bp in length
          *   20836   20935: gap of 100 bp
          *   20936   21631: contig of 696 bp in length
          *   21632   21731: gap of 100 bp
          *   21732   22419: contig of 688 bp in length
          *   22420   22519: gap of 100 bp
          *   22520   23215: contig of 696 bp in length
          *   23216   23315: gap of 100 bp
          *   23316   24016: contig of 701 bp in length
          *   24017   24116: gap of 100 bp
          *   24117   24826: contig of 710 bp in length
          *   24827   24926: gap of 100 bp
          *   24927   25637: contig of 711 bp in length
          *   25638   25737: gap of 100 bp
          *   25738   26460: contig of 723 bp in length
          *   26461   26560: gap of 100 bp
          *   26561   27261: contig of 701 bp in length
          *   27262   27361: gap of 100 bp
          *   27362   28075: contig of 714 bp in length
          *   28076   28175: gap of 100 bp
          *   28176   28880: contig of 705 bp in length
          *   28881   28980: gap of 100 bp
          *   28981   29685: contig of 705 bp in length
          *   29686   29785: gap of 100 bp
          *   29786   30479: contig of 694 bp in length
          *   30480   30579: gap of 100 bp
          *   30580   31254: contig of 675 bp in length
          *   31255   31354: gap of 100 bp
          *   31355   32042: contig of 688 bp in length
          *   32043   32142: gap of 100 bp
          *   32143   32847: contig of 705 bp in length
```

```
Db  27246  GTCTTTGCTATTTGAGATTGTGACCCACACAGGCGGTTGGCTGAAAGGGAGGCAGAAGGG  27187
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   5101  CGGGAGGGAGGGAATAGATGAAAAACAAGATGAAAAACAAACAAACTTCCCTAAGCAGTCT  5160
                 |||  | |||                  ||||||||||||||||||||||||||||||||
Db  27186  GGTGGGGGTGGGG------CAGGGAGGAGCCCAGAAAAAAAAAAAGTCCTAAGCGTCTCT  27132
                                    |||||||||  |  |||||||||||||||||||
Qy   5161  ACAAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGA  5220
                 |||||||||||||  |||||||| ||||||||  |||||||||||||||||||||||||
Db  27131  GCAAAACATTGTAGCCACAGGACTCCAAATCAGTGACAGAGAGCTCTCAAATCAGAATACAGA  27072
                 |||||||| ||| ||| |||  | |||||||   | |||||||||||||||||||||||
Qy   5221  TACAAGGAAGTGTATGTAGCTGGACAGGGTGGACACTCAGTTCAGTTCAGTTACA  5280
                 |||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  27071  TACAAGGAGATGTTATTCAGTTGGAGCAAGGGGACCATTTATTAGCTCAGT-------GA  27019
                 ||||||||| ||| ||||||||||  |||       |||| ||||||||||       ||
Qy   5281  AAAGTCCAGGCTGCTGAAATAAACCTCTGATAGTCCATTCATGCCAGCATCCAATCACAGACA  5340
                 |||||||||||||||||||||| |  ||||||||||||||||| |||||||||||||||
Db  27018  CAAGTCCTGGCTTCTGTGATTAAACTCTGATGCCATTCATCATACCAGCACCCAAGC  26959
                 ||||||| |||| || |||||  |||||||  |||||||||| ||||||| |||||
Qy   5341  GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAACAACAAGTGTGGCTACTA  5400
                 |||||||||||||||||||||||||||  ||||||  |||||||||||||||
Db  26958  AAGATCAGAAGTTCAGAGATGCCTACAAAT-------TGCCAACAAGTGTGGCCACTC  26908
                 |||||||||||||||||||||||        |||||||||||||||| |||||
Qy   5401  TACGTCAAGGACTCGAAGCCGTGAGAGAGGGGGAGAACACAGTAGAGGATGCCCA  5460
                 |||||||||||||||||||||||||||  |||| || ||||||||||||||||||
Db  26907  TACGTCAAGGGCTCTAAAACTGTGGCAGA-GAGGAAGAACAAGCGTTTACAGGGGGTGCCCA  26849
                 |||||||||  |||| ||| |||| || | |||||||||||  || ||||||||||||||
Qy   5461  GCTGGTAAGAATCGAGTGTTTATGAGAGTTTAGTCAATTCATTGGCTAAAA  5520
                 ||||||||| |||| |||||||||||| ||||||| |||||||||||||||
Db  26848  GCTGGTAAGAATTGACGGTTTTATGATGCTCTGGTTACTTGGAGACTTCATTGGCTGAAA  26789
                 |||||||| | || ||  ||||| || |||||||| ||| ||||||||||||||||||
Qy   5521  TCAAGAAACGCTCCGCCTCTTGCAAATATGTGTCAAATATGTATGAAGGAGGAAGTGCTAAACTTCTA  5580
                 |||||||| ||| ||||||| |||||||||||||||||||||||||||||||||||
Db  26788  GGAGAAACGCCCCGCCTCTTTGCAAATCTGAGTAAAGGGGGGGAAGTGTCTAAACTTCTA  26729
                 |||||| | ||| |||||||||||| ||| ||||||||||||||||||||||||||
Qy   5581  TGTCTGATAGAGCATTGACCCTATTGCTTTATTATCATTATAATATATACAC  5640
                 |||||||||||  ||||||||||||||||| |||| ||||||||||||||||
Db  26728  TGTCTGATGGACATTTGACCCTATTGCTTCAGCCTCCTGGCTACATACCTAGATATTCTC  26669
                 |||||||| |  || |||||||||||||   ||||||||| ||||| || ||||
Qy   5641  AGGTATTTGTGTATATTTTATATAATGTCTCCGT  5676
                 ||||| ||||| || ||||||| |||  ||||||
Db  26668  AGGTGTATATGTATATTTTTATAGAATTGCTTCCCAT  26633
```

```
RESULT 12
HSP63G01
LOCUS       HSP63G01                  835 bp    DNA     linear    PRI 04-JAN-2001
DEFINITION  Homo sapiens P63 protein (P63) gene, exon 1.
ACCESSION   AF124528
VERSION     AF124528.1  GI:12024731
KEYWORDS    .
SEGMENT     1 of 13
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 835)
AUTHORS     Yang,A., Kaghad,M., Wang,Y., Gillett,E., Fleming,M.D., Dotsch,V.,
            Andrews,N.C., Caput,D. and McKeon,F.
TITLE       p63, a p53 homolog at 3q27-29, encodes multiple products with
            transactivating, death-inducing, and dominant-negative activities
JOURNAL     Mol. Cell 2 (3), 305-316 (1998)
MEDLINE     98448095
PUBMED      9774969
REFERENCE   2  (bases 1 to 835)
AUTHORS     Hagiwara,K., McMenamin,M.G. and Harris,C.C.
TITLE       Direct Submission
JOURNAL     Submitted (29-JAN-1999) Laboratory of Human Carcinogenesis,
            National Cancer Institute, Building 37, Room 2C22, 37 Convent
            Drive, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1..835
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
*   32848    32947:  gap of 100 bp in length
*   33653:  contig of 706 bp in length
*   33754    33753:  gap of 100 bp in length
*   34469:  contig of 716 bp in length
*   34470    34569:  gap of 100 bp in length
*   35281:  contig of 712 bp in length
*   35282    35381:  gap of 100 bp in length
*   36076:  contig of 695 bp in length
*   36077    36176:  gap of 100 bp in length
*   36887:  contig of 711 bp in length
*   36888    36987:  gap of 100 bp in length
*   37684:  contig of 697 bp in length
*   37685    37784:  gap of 100 bp in length
*   38484:  contig of 700 bp in length
*   38485    38584:  gap of 100 bp in length
*   39274:  contig of 690 bp in length
*   39275    39374:  gap of 100 bp in length
*   40069:  contig of 695 bp in length
*   40070    40169:  gap of 100 bp in length
*   40980:  contig of 711 bp in length
*   40981    41700:  contig of 720 bp in length
*   41800:  gap of 100 bp in length
*   41801    42513:  contig of 713 bp in length
*   42514    42613:  gap of 100 bp in length
*   43309:  contig of 696 bp in length
*   43310    43409:  gap of 100 bp in length
*   44112:  contig of 703 bp in length
*   44113    44212:  gap of 100 bp in length
*   44914:  contig of 702 bp in length
*   44915    45014:  gap of 100 bp in length
*   45722:  contig of 708 bp in length
*   45723    45822:  gap of 100 bp in length
*   46541:  contig of 719 bp in length
*   46542    46641:  gap of 100 bp in length
*   47355:  contig of 714 bp in length
*   47356    47455:  gap of 100 bp in length
*   48171:  contig of 716 bp in length
*   48172    48271:  gap of 100 bp in length
*   48954:  contig of 683 bp in length
*   48955    49054:  gap of 100 bp in length
*   49774:  contig of 733 bp in length
*   49775    49874:  gap of 100 bp in length
*   50604:  contig of 730 bp in length
*   50605    50704:  gap of 100 bp in length
*   51409:  contig of 705 bp in length
*   51410    51509:  gap of 100 bp in length
*   52214:  contig of 705 bp in length
*   52215    52314:  gap of 100 bp in length
*   52991:  contig of 677 bp in length
*   52992    53091:  gap of 100 bp in length
*   53795:  contig of 704 bp in length
*   53796    53895:  gap of 100 bp in length
*   54628:  contig of 733 bp in length
*   54629    54728:  gap of 100 bp in length
*   55450:  contig of 722 bp in length
*   55451    55550:  gap of 100 bp in length
*   56268:  contig of 718 bp in length.
```

```
Query Match    6.6%;  Score 372.6;  DB 2;  Length 56268;
Best Local Similarity  67.3%;  Pred. No. 4.5e-61;
Matches  509;  Conservative  0;  Mismatches  225;  Indels  22;  Gaps  4;

Qy   4921  AGGTTATTTCTTTATTCGGGTCAGCAAAGCTTCTAAGGGGATGTGAAAGGGATATCT  4980
                 |||
Db  27366  AGGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  27307
Qy   4981  CTTTCTCTTAGCTGAGGAAGAGTGAGTTCTAAGTTAAATATAATAATCAAGGAAATTTCCCT  5040
                 ||||||||||
Db  27306  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACAAGGAATTTCCCT  27247
Qy   5041  GTCTTTGCTATTTGAGATTGTGACCCACACAGGCGGTTGGCTGAAAGGGAGG  5100
                 ||||||
```

```
        exon           /db_xref="taxon:9606"
                       216..305
                       /gene="P63"
                       /number=1
ORIGIN

Query Match            3.8%;    Score 215;  DB 9;  Length 835;
Best Local Similarity  100.0%;  Pred. No. 8e-31;
Matches  215;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    5462  CTGGTAAGAATCGAGTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAAT  5521
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  CTGGTAAGAATCGAGTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAAT    60

Qy    5522  CAAGAAACGCTCCGGCCTCTTTGCAAATATGTATGAAGGAGGAGAAGTGCCTAAACTTCTAT  5581
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  CAAGAAACGCTCCGGCCTCTTTGCAAATATGTATGAAGGAGGAGAAGTGCCTAAACTTCTAT   120

Qy    5582  GTCTGATAGCATTGACCCTATTGCTTTAGCCTCCGGCTTTATCTATATATACACA  5641
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  GTCTGATAGCATTGACCCTATTGCTTTAGCCTCCGGCTTTATCTATATATACACA   180

Qy    5642  GGTATTTGTGTATATTTTATATAATTGTTCTCCGT  5676
            |||||||||||||||||||||||||||||||||||
Db     181  GGTATTTGTGTATATTTTATATAATTGTTCTCCGT   215
```

```
RESULT 13
AC103882/c
LOCUS       AC103882         159249 bp    DNA     linear    PRI 15-OCT-2002
DEFINITION  Homo sapiens BAC clone RP11-733G6 from 2, complete sequence.
ACCESSION   AC103882
VERSION     AC103882.5  GI:23130727
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159249)
AUTHORS     Sulston,J.E. and Waterston,R.
TITLE       Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
PUBMED      9847074
REFERENCE   2  (bases 1 to 159249)
AUTHORS     Wang,C. and Bielicki,L.
TITLE       The sequence of Homo sapiens BAC clone RP11-733G6
JOURNAL     Unpublished (2001)
REFERENCE   3  (bases 1 to 159249)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (29-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159249)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (30-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159249)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (18-SEP-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 159249)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-OCT-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 159249)
AUTHORS     Waterston,R.
```

```
TITLE       Direct Submission
JOURNAL     Submitted (15-OCT-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Sep 18, 2002 this sequence version replaced gi:19703341.
COMMENT     -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
            Center project name: H_NH0733G06
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR: pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is not the entire insert of the clone. This clone is
            overlapped by AC131754 and AC096656.

            There is a simple sequence repeat from base 46448 to 46739. This
            region does not meet required finishing standards.
FEATURES             Location/Qualifiers
     source          1..159249
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-733G6"
                     /clone_lib="RPCI-11"
     repeat_region   102..178
                     /rpt_family="L1"
     repeat_region   179..571
                     /rpt_family="MaLR"
     repeat_region   572..626
                     /rpt_family="L1"
     repeat_region   765..942
                     /rpt_family="L1"
     repeat_region   1024..1916
                     /rpt_family="L1"
     repeat_region   1917..2217
                     /rpt_family="Alu"
     repeat_region   2218..2499
                     /rpt_family="L1"
     repeat_region   2499..3481
                     /rpt_family="L1"
```

```
repeat_region   3482..3782
                /rpt_family="Alu"
repeat_region   3783..3929
                /rpt_family="L1"
repeat_region   3930..4251
                /rpt_family="L1"
repeat_region   4252..4764
                /rpt_family="Alu"
repeat_region   4776..4986
                /rpt_family="L1"
repeat_region   4994..5023
                /rpt_family="L1"
repeat_region   5024..5176
                /rpt_family="(TA)n"
repeat_region   5236..5404
                /rpt_family="L1"
repeat_region   5520..5705
                /rpt_family="L1"
repeat_region   5747..5777
                /rpt_family="L1"
repeat_region   5782..5972
                /rpt_family="AT_rich"
repeat_region   6450..6474
                /rpt_family="L1"
repeat_region   6998..7244
                /rpt_family="(TG)n"
repeat_region   7948..8299
                /rpt_family="MIR"
repeat_region   8868..9128
                /rpt_family="MER2_type"
repeat_region   9299..9510
                /rpt_family="MER2_type"
repeat_region   9516..9711
                /rpt_family="ERVL"
repeat_region   9717..10011
                /rpt_family="ERVL"
repeat_region   10275..10313
                /rpt_family="Alu"
repeat_region   10375..10528
                /rpt_family="AT_rich"
repeat_region   10529..10659
                /rpt_family="(TA)n"
repeat_region   10696..11412
                /rpt_family="(TATAA)n"
repeat_region   11440..11544
                /rpt_family="ERVL"
repeat_region   11569..11927
                /rpt_family="ERV1"
repeat_region   12292..12495
                /rpt_family="MaLR"
repeat_region   12856..13041
                /rpt_family="ERV1"
repeat_region   13197..14809
                /rpt_family="ERVL"
repeat_region   14829..14876
                /rpt_family="L1"
repeat_region   14998..15187
                /rpt_family="MIR"
repeat_region   15243..15523
                /rpt_family="MER1_type"
repeat_region   15728..15999
                /rpt_family="Alu"
repeat_region   16186..16418
                /rpt_family="Alu"
repeat_region   16727..17022
                /rpt_family="MaLR"
repeat_region   17166..17468
                /rpt_family="Alu"
repeat_region   17670..17868
                /rpt_family="Alu"
repeat_region   18530..18674
                /rpt_family="MIR"
repeat_region   19048..19069
                /rpt_family="MIR"
```

```
repeat_region   20886..21207
                /rpt_family="AT_rich"
repeat_region   21379..22209
                /rpt_family="CR1"
repeat_region   22342..22558
                /rpt_family="L1"
repeat_region   22572..22872
                /rpt_family="MIR"
repeat_region   23219..23264
                /rpt_family="CR1"
```

```
Query Match     3.7%;   Score 212.4;  DB 9;   Length 159249;
Best Local Similarity   63.3%;  Pred. No. 1.6e-30;
Matches 413;    Conservative   0;   Mismatches 221;   Indels  18;   Gaps   5;

Qy     3293 TGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAGGCCGGGTCGGGAGGATAGCTTGAG 3352
            ||||||||||| ||||||| ||||||||||||||||| ||| ||||||||||| ||||||
Db    85766 TGGTGGCTTCTGCCTGAACTCCCAGCACTTTGGGAGGCCAAGGTGGGAGGATCCTTTGAG 85707

Qy     3353 TCCAGCAGTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAACAAAA 3412
            ||||||||| |||| |||| |||||| ||| ||||| |||| | || |||||||| ||||
Db    85706 CTCAGGCATTTGAGACCAGCCAGGCCTGGGCAATATAAAAAGGTCTCGTGTCTACTAAAATTACA 85647

Qy     3413 CAAAACAAAATTACCTGGGTATGGTGGTGTGCTCACCTGTAGTCCAAGCTACACAGGAGCT 3472
            ||||||||||        |||||||||||||||||||||| ||||| ||| ||||||| |
Db    85646 ------AAAATTAGCTGGGTGTGGTGTGTACATGCCTGTAGTCCCAGCTACTCAGGAGGCT 85593

Qy     3473 GAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGCTGCAGTGATCCATGAACGCCCTG 3532
            |||||| ||||||||||||||||||||||||||| |||||||||||| |||| || ||||
Db    85592 GAAGCCGGAAGGATCACCTGAGCCCAGGAGGTAGAGGCTGCAGTGAAGATGGCACCA 85533

Qy     3533 CTACACT-CAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAATAAATAAAA 3591
            ||| |||  ||||| ||||| || |||||| ||||| ||| |||||| || |||||||||
Db    85532 CTGTACTCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAAG 85473

Qy     3592 ATAACTTTTAAAAACAAAATTAATTAAATTTAAAAACAACACCACTAGAGAT--GT 3649
            |||| ||| |||||||||| ||||||||||  |||| ||| ||||| |||| || |
Db    85472 AAAAGAAAAGAAAAAAGAAAGAAAGAAAAAAACAACAAGAAAAACAATGGTTAAA 85413

Qy     3650 TTGCAAATTGATTATTTGGGAGTCTATATCCCTGGAAGTTAATTTAAAATATTTAGAAGA 3709
            ||||||   |||| ||||| || |||||||||| || |||| |||||||||||||||||||
Db    85412 AAGCAAATTTGTGCCTGGGAAAATCACTAATATCCAGTGTTCTGTTATCACATAGCAGA 85353

Qy     3710 GTTCTTCCTCATTTCCTAGAGACGTCGAATTGTAAATATCAGAGCTAGAAGGAACACTAG 3769
            |||||||| ||||  |||||||| || |||||  | ||| || ||||||||| |||  ||
Db    85352 AAACAAGCCCTTCAGTTAGCGAAACTCGGGTTGAGGGAGAGCAAGGCAAGCAGCCAAGAA 85293

Qy     3770 GGCTCGCCACTCCAAAGTGTGTCCAAGACCAGCAGCATCAAGTAACCTGGGAACGTGT 3829
            || |   ||||  |||||||| || ||||||                           ||
Db    85292 GGTGCTAGGGAAAGAGAACTGGAGAATGGTGCAGGGCAC------TGCCTGGGAGCTTGT 85240

Qy     3830 TAGAAATGCAGAGTCTTAGGGCTCACCCCAGACCTCACCCCAGACCA--GAATCTGCATTAAC 3887
            |||||||||||||||||||| ||||||||| ||||| ||| |||| |||| ||||| ||||
Db    85239 CAGAAATGCAGAGTCTTGGGCTTGGGCTTCTCCCCAGACCCCACTGAATCAGGGCATGTATTTTAAC 85180

Qy     3888 AAGATTTCTAGGTGCCTCACGGCACATTAAAACTTGAGAAGTCTGCACTA 3939
            ||||  ||||||||||  |||||||| ||||||||||||||||| |||||
Db    85179 AAGACCCTCAGGTGATTCCTGCACAGGTTAAAACTTGAAGAAGCACTCGCACAA 85128
```

```
RESULT 14
AC087334
LOCUS         AC087334               60323 bp    DNA     linear   HTG 30-DEC-2001
DEFINITION    Homo sapiens chromosome 17 clone RP11-73H6 map 17, LOW-PASS
              SEQUENCE SAMPLING.
ACCESSION     AC087334
VERSION       AC087334.3  GI:17798705
KEYWORDS      HTG; HTGS PHASE0.
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 60323)
```

AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome 17, clone RP11-73H6
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 60323)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
           Canarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
           Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
           Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
           Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
           Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
           McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
           O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhagg,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
           Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
           Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
           Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (28-DEC-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Dec 30, 2001 this sequence version replaced gi:12658045.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
           ----------------- Project Information
              Center project name: L10827
              Center clone name: 73_H_6
           ----
         * NOTE: This record contains 72 individual
         * sequencing reads that have not been assembled into
         * contigs. Runs of N are used to separate the reads
         * and the order in which they appear is completely
         * arbitrary. Low-pass sequence sampling is useful for
         * identifying clones that may be gene-rich and allows
         * overlap relationships among clones to be deduced.
         * However, it should not be assumed that this clone
         * will be sequenced to completion. In the event that
         * the record is updated, the accession number will
         * be preserved.
         *   1   722: contig of 722 bp in length
         *   723   822: gap of 100 bp
         *   823   1575: contig of 753 bp in length
         *   1576   1675: gap of 100 bp
         *   1676   2434: contig of 759 bp in length
         *   2435   2534: gap of 100 bp
         *   2535   3262: contig of 728 bp in length
         *   3263   3362: gap of 100 bp
         *   3363   4097: contig of 735 bp in length
         *   4098   4197: gap of 100 bp
         *   4198   4951: contig of 754 bp in length
         *   4952   5051: gap of 100 bp
         *   5052   5810: contig of 759 bp in length
         *   5811   5910: gap of 100 bp
         *   5911   6680: contig of 770 bp in length
         *   6681   6780: gap of 100 bp
         *   6781   7511: contig of 731 bp in length
         *   7512   7611: gap of 100 bp
         *   7612   8341: contig of 730 bp in length
         *   8342   8441: gap of 100 bp
         *   8442   9174: contig of 733 bp in length

         *   9175   9274: gap of 100 bp
         *   9275   10016: contig of 742 bp in length
         *   10017   10116: gap of 100 bp
         *   10117   10860: contig of 744 bp in length
         *   10861   10960: gap of 100 bp
         *   10961   11709: contig of 749 bp in length
         *   11710   11809: gap of 100 bp
         *   11810   12383: contig of 574 bp in length
         *   12384   12483: gap of 100 bp
         *   12484   13244: contig of 761 bp in length
         *   13245   13344: gap of 100 bp
         *   13345   14041: contig of 697 bp in length
         *   14042   14141: gap of 100 bp
         *   14142   14894: contig of 753 bp in length
         *   14895   14994: gap of 100 bp
         *   14995   15751: contig of 757 bp in length
         *   15752   15851: gap of 100 bp
         *   15852   16575: contig of 724 bp in length
         *   16576   16675: gap of 100 bp
         *   16676   17401: contig of 726 bp in length
         *   17402   17501: gap of 100 bp
         *   17502   18237: contig of 736 bp in length
         *   18238   18337: gap of 100 bp
         *   18338   19073: contig of 736 bp in length
         *   19074   19173: gap of 100 bp
         *   19174   19907: contig of 734 bp in length
         *   19908   20007: gap of 100 bp
         *   20008   20762: contig of 755 bp in length
         *   20763   20862: gap of 100 bp
         *   20863   21569: contig of 707 bp in length
         *   21570   21669: gap of 100 bp
         *   21670   22380: contig of 711 bp in length
         *   22381   22480: gap of 100 bp
         *   22481   23231: contig of 751 bp in length
         *   23232   23331: gap of 100 bp
         *   23332   24101: contig of 770 bp in length
         *   24102   24201: gap of 100 bp
         *   24202   24959: contig of 758 bp in length
         *   24960   25059: gap of 100 bp
         *   25060   25780: contig of 721 bp in length
         *   25781   25880: gap of 100 bp
         *   25881   26614: contig of 734 bp in length
         *   26615   26714: gap of 100 bp
         *   26715   27436: contig of 722 bp in length
         *   27437   27536: gap of 100 bp
         *   27537   28277: contig of 741 bp in length
         *   28278   28377: gap of 100 bp
         *   28378   29108: contig of 731 bp in length
         *   29109   29208: gap of 100 bp
         *   29209   29963: contig of 755 bp in length
         *   29964   30063: gap of 100 bp
         *   30064   30831: contig of 768 bp in length
         *   30832   30931: gap of 100 bp
         *   30932   31641: contig of 710 bp in length
         *   31642   31741: gap of 100 bp
         *   31742   32490: contig of 749 bp in length
         *   32491   32590: gap of 100 bp
         *   32591   33338: contig of 748 bp in length
         *   33339   33438: gap of 100 bp
         *   33439   34170: contig of 732 bp in length
         *   34171   34270: gap of 100 bp
         *   34271   35010: contig of 740 bp in length
         *   35011   35110: gap of 100 bp
         *   35111   35936: contig of 826 bp in length
         *   35937   36036: gap of 100 bp
         *   36037   36800: contig of 764 bp in length
         *   36801   36900: gap of 100 bp
         *   36901   37632: contig of 732 bp in length
         *   37633   37732: gap of 100 bp
         *   37733   38436: contig of 704 bp in length
         *   38437   38536: gap of 100 bp
         *   38537   39309: contig of 773 bp in length
         *   39310   39409: gap of 100 bp

```
Db    9995  TAATTAATTAAAAAAAGAAAAAANNNNNNNNNNNNNNNNNNNN  10054
                 |||  ||| ||||  |
Qy    3683  GGAAGTTAATTTAAAATATTTAGAAGAGTTCTTCCTCCATTCCTAGAGACGTCGAATTGT  3742

Db   10055  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  10114
                 |||| |
Qy    3743  AAATATCAGAGCT  3755
                 ||| ||
Db   10115  NNAGGCGGAGATCT  10127
```

```
RESULT 15
AL158198
LOCUS         AL158198           173456 bp    DNA      linear    PRI 26-OCT-2000
DEFINITION    Human DNA sequence from clone RP11-239H6 on chromosome 6 Contains
              an STS, GSSs and CpG islands, complete sequence.
ACCESSION     AL158198
VERSION       AL158198.14  GI:9944140
KEYWORDS      HTG; CpG island.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 173456)
AUTHORS       Tracey,A.
TITLE         Direct Submission
JOURNAL       Submitted (20-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
              requests: clonerequest@sanger.ac.uk
COMMENT       On Aug 29, 2000 this sequence version replaced gi:9843926.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone contigs of human
              chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
              Group.  Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr6
              RP11-239H6 is from the library RPCI-11.1 constructed at the Roswell
              Park Cancer Institute by the group of Pieter de Jong. For further
              details see http://bacpac.med.buffalo.edu/
              VECTOR: pBACe3.6
              IMPORTANT: This sequence is not the entire insert of clone
              RP11-239H6 It may be shorter because we sequence overlapping
              sections only once, except for a 100 base overlap.
              The true left end of clone RP11-239H6 is at 1 in this sequence. The
              true left end of clone RP1-46C2 is at 173357 in this sequence. The
              true right end of clone RP3-434O1 is at 62736 in this sequence.
FEATURES      Location/Qualifiers
     source   1..173456
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /chromosome="6"
              /clone="RP11-239H6"
              /clone_lib="RPCI-11.1"
     repeat_region  1609..2046
              /note="LTR16B repeat: matches 1..464 of consensus"
     repeat_region  2688..2713
```

```
*  39410
*  40153:  contig of 744 bp in length
*  40154
*  40253:  gap of 100 bp
*  40981:  contig of 728 bp in length
*  40982
*  41081:  gap of 100 bp
*  41082
*  41819:  contig of 738 bp in length
*  41820
*  41919:  gap of 100 bp
*  41920
*  42646:  contig of 727 bp in length
*  42647
*  42746:  gap of 100 bp
*  42747
*  43477:  contig of 731 bp in length
*  43478
*  43577:  gap of 100 bp
*  43578
*  44310:  contig of 733 bp in length
*  44311
*  44410:  gap of 100 bp
*  44411
*  45176:  contig of 766 bp in length
*  45177
*  45276:  gap of 100 bp
*  45277
*  46043:  contig of 767 bp in length
*  46044
*  46143:  gap of 100 bp
*  46144
*  46849:  contig of 706 bp in length
*  46850
*  46949:  gap of 100 bp
*  46950
*  47701:  contig of 752 bp in length
*  47702
*  47801:  gap of 100 bp
*  47802
*  48543:  contig of 742 bp in length
*  48544
*  48643:  gap of 100 bp
*  48644
*  49411:  contig of 768 bp in length
*  49412
*  49511:  gap of 100 bp
*  49512
*  50236:  contig of 725 bp in length
*  50237
*  50336:  gap of 100 bp
*  50337
*  51067:  contig of 731 bp in length
*  51068
*  51167:  gap of 100 bp
*  51168
*  51895:  contig of 728 bp in length
*  51896
*  51995:  gap of 100 bp
*  51996
*  52703:  contig of 708 bp in length
*  52704
*  52803:  gap of 100 bp
*  52804
*  53550:  contig of 747 bp in length
*  53551
*  53650:  gap of 100 bp
*  53651
*  54364:  contig of 714 bp in length
*  54365
*  54464:  gap of 100 bp
*  54465
*  55227:  contig of 763 bp in length
*  55228
*  55327:  gap of 100 bp
*  55328
*  56118:  contig of 791 bp in length
*  56119
*  56218:  gap of 100 bp
*  56219
*  56961:  contig of 743 bp in length
*  56962
*  57061:  gap of 100 bp
*  57062
*  57773:  contig of 712 bp in length
```

```
Query Match    3.6%;   Score 205;  DB 2;  Length 60323;
Best Local Similarity  57.2%;  Pred. No. 4.4e-29;
Matches  282;  Conservative  0;  Mismatches 210;  Indels  1;  Gaps  1;

Qy   3264  TTTAAATTTTTTAAAAAAAATAGATGAGTGTGCTGTGGCTCATGCTGTAATCCCAACACTTT  3323
                |||| ||||  |||||  | |||||||||||||||||||||||| ||||||
Db   9635  TCTAATTTTAAGAAAAAAAATGGCTTGGGGATGGTGGCTCATGCTGTAATCCTAACACTTT  9694

Qy   3324  GGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACACAGTCAGGGCAACAC  3383
                |||||||| || | ||||||||||||||||||||||||||| ||||||||||||||||
Db   9695  GGGAGGCCAAGGCGGATGGATCGCCTCAGCTCAGGCGTTTGCAACCAGCCTGGGCAACAC  9754

Qy   3384  AGCAAGACCCCATATCTAAAAACAAAACAAAACAAAACAAAAATTACCTGGGTATGGTTCTGCT  3443
                || ||||| |||||  ||||  ||||||||||||| ||||||||||||
Db   9755  GGTGAAACCCCGTCTCTACTAAAAAAATACAAAAAAATTACCTGGGCATGGTGATGTG  9814

Qy   3444  CACCTGTAGTCCAAGCTACACAGGAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGT  3503
                |||||||| || ||| |||||  ||||| ||||||||||||||| ||||||||||||||
Db   9815  CACCTGTAATCCCAGCACTCAAGAGGCTGAGGCTGGGAGGATCACTTGAACCCAGGAGGC  9874

Qy   3504  TGAGGCTGCAGTGATCCATGAACGCGCCGGCTCGCTACACT-CAGTCTCGGGTGACAGTGCAAGAA  3562
                | ||||||||||||  ||||   || || |||||||||  ||| |  ||||  || ||
Db   9875  GGAGGTTGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGAC  9934

Qy   3563  GCTGTCTTCAAAATAATATAAATAAATAAAACAAAATAACTTTTAAAAAACAAAAATTAATTAAAT  3622
                ||| | | |||||  |||  | |||||||||||||| |||  |||||||||||||
Db   9935  TCCGTCTCAAAAAATAAAAATAAATAAAAAATAAATAAATAATTAATTAATTAATTAAT  9994

Qy   3623  TTTAAAAACACACACACTAGAGATGTTGCAAATTGATTATTGGGAGTCTATATCCCT  3682
```

repeat_region   /note="13 copies 2 mer tt 92% conserved"
4959. .5048
repeat_region   /note="L1M1 repeat: matches 5492. .5581 of consensus"
5508. .5533
misc_feature   /note="13 copies 2 mer ac 100% conserved"
complement(5984. .6363)
repeat_region   /note="match: GSS: Em:AQ805521"
8132. .8622
misc_feature   /note="MLT2CB repeat: matches 1. .501 of consensus"
11067. .11517
misc_feature   /note="match: GSS: Em:AQ886344"
11303. .11541
repeat_region   /note="L1PA13 repeat: matches 5881. .6140 of consensus"
complement(11581. .12028)
misc_feature   /note="match: GSS: Em:AQ464397"
complement(11631. .12028)
misc_feature   /note="match: GSS: Em:AQ488628"
12033. .12675
misc_feature   /note="match: GSS: Em:AQ239805"
12035. .12449
misc_feature   /note="match: GSS: Em:AQ828417"
12037. .12660
misc_feature   /note="match: GSS: Em:AQ236775"
12041. .12600
misc_feature   /note="match: GSS: Em:AQ489650"
12056. .12447
misc_feature   /note="match: GSS: Em:AQ045010"
12299. .12387
repeat_region   /note="MER5B repeat: matches 83. .178 of consensus"
12875. .13212
repeat_region   /note="MLT1A1 repeat: matches 11. .365 of consensus"
14682. .15037
repeat_region   /note="THE1B repeat: matches 1. .364 of consensus"
15442. .15516
repeat_region   /note="L1ME3 repeat: matches 6063. .6116 of consensus"
18055. .18255
repeat_region   /note="L1ME1 repeat: matches 5930. .6121 of consensus"
complement(19123. .19461)
misc_feature   /note="match: GSS: Em:AQ564070"
23382. .23776
repeat_region   /note="MER5B repeat: matches 7. .368 of consensus"
24079. .24106
repeat_region   /note="MER93 repeat: matches 368. .397 of consensus"
24500. .24637
repeat_region   /note="L1P repeat: matches 4241. .4385 of consensus"
24630. .24754
repeat_region   /note="L1P repeat: matches 5109. .5233 of consensus"
25993. .26923
repeat_region   /note="L1MA6 repeat: matches 4769. .5706 of consensus"
26933. .27145
repeat_region   /note="Charlie3 repeat: matches 1. .215 of consensus"
27146. .27585
repeat_region   /note="L1MA6 repeat: matches 5719. .6300 of consensus"
28238. .28386
repeat_region   /note="L1 repeat: matches 5151. .5308 of consensus"
28765. .29297
repeat_region   /note="MER1A repeat: matches 1. .527 of consensus"
29622. .30025
misc_feature   /note="MLT1B repeat: matches 1. .390 of consensus"
complement(31073. .31648)
misc_feature   /note="match: GSS: Em:AQ345866"
31965. .32002
repeat_region   /note="19 copies 2 mer ac 97% conserved"
32293. .32487
repeat_region   /note="MER5B repeat: matches 7. .176 of consensus"
32550. .32908
repeat_region   /note="MER47A repeat: matches 1. .366 of consensus"
35097. .35318
repeat_region   /note="L1ME repeat: matches 5371. .5603 of consensus"
36107. .36248
repeat_region   /note="AluSg/x repeat: matches 133. .274 of consensus"
36249. .36614
/note="183 copies 2 mer ga 66% conserved"

repeat_region   37398. .37431
/note="17 copies 2 mer aa 85% conserved"
repeat_region   40629. .41044
/note="MSTA repeat: matches 1. .426 of consensus"
repeat_region   42019. .42574
/note="MER7C repeat: matches 1. .602 of consensus"
misc_feature   46407. .47179
/note="match: GSS: Em:AQ385283"
46739. .46926
repeat_region   /note="L1PA10 repeat: matches 5958. .6153 of consensus"
47112. .47740
repeat_region   /note="L1PA10 repeat: matches 5365. .5987 of consensus"
49919. .50232
repeat_region   /note="MLT1A1 repeat: matches 1. .374 of consensus"
51550. .51903
misc_feature   /note="THE1A repeat: matches 1. .354 of consensus"
complement(52137. .52547)
repeat_region   52671. .52702
/note="match: GSS: Em:AQ016629"
repeat_region   53069. .53456
/note="16 copies 2 mer gt 100% conserved"
misc_feature   /note="MLT2FB repeat: matches 1. .412 of consensus"
complement(56179. .56798)
repeat_region   56189. .56750
/note="match: GSS: Em:AQ267603"
misc_feature   56796. .57235
/note="MER61E repeat: matches 1. .569 of consensus"
repeat_region   57071. .57187
/note="match: GSS: Em:AQ137177"
repeat_region   57354. .57759
/note="L1MA8 repeat: matches 6123. .6238 of consensus"
repeat_region   58839. .59181
/note="L1MA8 repeat: matches 5713. .6123 of consensus"
repeat_region   61949. .62748
/note="MSTD repeat: matches 1. .394 of consensus"
repeat_region   63038. .63383
/note="MER61E repeat: matches 6961. .7734 of consensus"
misc_feature   /note="L1MC3 repeat: matches 6632. .6961 of consensus"
complement(63091. .63520)
misc_feature   63188. .63781
/note="match: GSS: Em:AQ818541"
repeat_region   63396. .63431
/note="match: GSS: Em:AQ485404"
repeat_region   65033. .65652
/note="18 copies 2 mer tg 94% conserved"
repeat_region   66656. .67052
/note="MER4A repeat: matches 2. .660 of consensus"
misc_feature   68264. .68774
/note="MER57B repeat: matches 1. .403 of consensus"
repeat_region   69704. .69789
/note="match: GSS: Em:AQ885163"
misc_feature   70719. .71158
/note="43 copies 2 mer aa 68% conserved"
repeat_region   71459. .71576
/note="match: GSS: Em:AQ805723"
repeat_region   75291. .75453
/note="MER57-internal repeat: matches 7151. .7270 of consensus"
repeat_region   77757. .77898
/note="MER58 repeat: matches 28. .196 of consensus"
repeat_region   79760. .79799
/note="LTR33 repeat: matches 364. .513 of consensus"
misc_feature   79828. .80692
/note="20 copies 2 mer ac 95% conserved"
repeat_region   81464. .81505
/note="CpG island"
/evidence=not_experimental
/note="21 copies 2 mer ga 76% conserved"

Query Match   3.6%;   Score 204.4;   DB 9;   Length 173456;
Best Local Similarity   72.3%;   Pred. No. 5.2e-29;
Matches 279;   Conservative   0;   Mismatches 106;   Indels   1;   Gaps   1;

```
Qy    3264  TTTAAATTTTTTAAAAAATAGATGAGTGTGGTGGCTCATGCCTGTAATCCCAACACTTT  3323
                ||||| ||||| ||| ||| |||||||||||||||||||||| |||||||
Db   19809  TCTAATTTTAAGAAAAAATGGGCTGGGGATGGTGCTCATGCCTGTAATCCTAACACTTT  19868

Qy    3324  GGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACAC  3383
                || |||| |||| ||||| ||| ||||| |||||| ||||||||||||||
Db   19869  GGGAGGCCAAGGCGGATGGATCGCCTGAGCTCAGGCGTTTGCAACCAGCCTGGGCAACAC  19928

Qy    3384  AGCAAGACCCCATATCTARAAAAACAAAACAAAATTACCTGGGTATGGTTGTGCT  3443
                 |||||| ||||| |||||||     |||||||||  ||||| |||
Db   19929  GGTGAAACCCCGTCTCTACTAAAAAAATACAAAAAAATTACCTGGGCATGGTGATGTG  19988

Qy    3444  CACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGAATCACTTGAGCCCAGGAGGT  3503
                |||||| || |||| ||| |||||||||| ||||||||||||
Db   19989  CACCTGTAATCCCAGCACTCAAGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGC  20048

Qy    3504  TGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAA  3562
                |||| |||||| |||| || ||||| |||| || || ||
Db   20049  GGAGGTCGCAGTGAGCCAAGATCGCCGCCACTGCGCCACTACAGCCTGGGCGACAGAGCGAGAC  20108

Qy    3563  GCTGTCTCAAAAATAATATAATTAAAATAACTTTTAAAAAACAAAAATTAATTAAAT  3622
                ||||||||| ||||  ||| |||  |||| ||||
Db   20109  TCCGTCTCAAAAAAAATAATAATAAATAAATAAATAAATAATTAATTAATTAAT  20168

Qy    3623  TTTAAAAACACAACCACACTAGAGATG  3648
                 || ||||| ||| |  ||| |
Db   20169  TAATTAATTAAAAAGAAAAAGTG  20194
```

Search completed: June 23, 2005, 07:51:11
Job time : 23476 secs

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   June 22, 2005, 14:13:22 ; Search time 2718 Seconds
                   (without alignments)
                   12362.197 Million cell updates/sec

Title:           US-10-030-294-1
Perfect score:   5676
Sequence:        1 cagctgttcaggggatgctg.........tttatataattgttcctccgt 5676

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:     8780412

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_16Dec04:*
              1: geneseqn1980s:*
              2: geneseqn1990s:*
              3: geneseqn2000s:*
              4: geneseqn2001as:*
              5: geneseqn2001bs:*
              6: geneseqn2002as:*
              7: geneseqn2002bs:*
              8: geneseqn2003as:*
              9: geneseqn2003bs:*
             10: geneseqn2003cs:*
             11: geneseqn2003ds:*
             12: geneseqn2004as:*
             13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5676	100.0	5676	5	AAF25904	Aaf25904 Human tum
c	2	5676	100.0	5676	5	AAF25907	Aaf25907 Human tum
c	3	5676	100.0	5676	5	AAF25906	Aaf25906 Human tum
	4	5676	100.0	5960	5	AAF25905	Aaf25905 Human tum
	5	5673.4	100.0	13940	5	AAF25915	Aaf25915 Human tum
c	6	196.2	3.5	58326	12	ADO48537	Ado48537 Human man
c	7	196	3.5	304905	10	ADP75180	Adp75180 Human End
c	8	193.4	3.4	118384	10	ABX56555	Abx56555 Human aut
	9	189.4	3.3	1384	3	AAC74312	Aac74312 Human sec
	10	188.4	3.3	135005	6	ADQ19501	Adq19501 Human sof
	11	187.6	3.3	8133	6	ABN99663	Abn99663 Human clu
	12	186	3.3	18861	12	ADJ12659	Adj12659 DNA fragm
	13	185.4	3.3	47999	8	AAD52898	Aad52898 Human twe
	14	184.8	3.3	199878	10	ADL13719	Adl13719 Osteoearth
c	15	183	3.2	335199	10	ADC24703	Adc24703 Human wil
c	16	182.6	3.2	47903	13	ABD33591	Abd33591 Human can
	17	182.2	3.2	36651	6	AAD28072	Aad28072 Human kin
c	18	181.8	3.2	49939	9	ADB16928	Adb16928 Human DYX
	19	181	3.2	58985	9	ABZ59738	Abz59738 Human sec
	20	181	3.2	143601	8	ABZ22654	Abz22654 Human epi
	21	180.8	3.2	198522	11	ACN44010	Acn44010 Human gen
c	22	180	3.2	26329	4	AAS42041	Aas42041 Genomic s
	23	180	3.2	94719	10	ADE95902	Ade95902 Human STA
	24	180	3.2	94720	9	ADA02654	Ada02654 Human STA
	25	180	3.2	94720	10	ADB72392	Adb72392 Human STA
	26	179.8	3.2	108316	10	ADC87336	Adc87336 Human GPC
c	27	179.6	3.2	65608	6	ABL62910	Abl62910 Breast ca
c	28	179.6	3.2	65608	6	ABL64414	Abl64414 Stomach c
c	29	179.6	3.2	65608	6	ABL67668	Abl67668 Oesophagu
c	30	179.4	3.2	214520	10	ADL13471	Adl13471 Osteoearth
	31	179.2	3.2	94330	11	ACN44662	Acn44662 Human gen
c	32	179.2	3.1	186957	8	AAD56071	Aad56071 Human sec
c	33	178.6	3.1	201986	11	ACN44430	Acn44430 Human gen
	34	178.4	3.1	47219	13	ADR16283	Adr16283 Human Cay
	35	178.4	3.1	59884	13	ADR16284	Adr16284 Human Cay
c	36	178.2	3.1	630		AAL15941	Aal15941 Human bre
c	37	178.2	3.1	906	11	ACN85902	Acn85902 Breast ca
	38	178.2	3.1	7726	5	ABA21079	Aba21079 Human ner
	39	178.2	3.1	25423	4	AAK90279	Aak90279 Human dig
	40	178.2	3.1	25423	4	AAI57656	Aai57656 Human col
	41	178.2	3.1	25423	6	ABS99833	Abs99833 Genomic D
	42	178.2	3.1	25423	10	ADB92986	Adb92986 Human col
	43	178.2	3.1	25424		AAK90280	Aak90280 Human dig
	44	178.2	3.1	25424	4	AAI57657	Aai57657 Human col
	45	178.2	3.1	25424	6	ABS99834	Abs99834 Genomic D

ALIGNMENTS

```
RESULT 1
AAF25904
ID   AAF25904 standard; DNA; 5676 BP.
XX
AC   AAF25904;
XX
DT   19-APR-2001  (first entry)
XX
DE   Human tumor suppressor gene p51 promoter associated DNA SEQ ID 1.
XX
KW   Tumor suppressor; p51; cell death; cell proliferation; cancer;
KW   cytostatic; gene therapy; screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100818-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-JP004261.
PR   29-JUN-1999;  99JP-00183195.
XX
PA   (NIPK ) NIPPON KAYAKU KK.
PA   (SAKA/) SAKAI T.
XX
PI   Sakai T, Kagaya S, Sato T, Sukenaga Y, Fujii H;
XX
DR   WPI; 2001-112452/12.
XX
PT   New DNA sequences encoding the human p51 promoter domain for screening
PT   for compounds that modify p51 promoter activity and for diagnosing and
PT   treating cancer.
XX
PS   Claim 1(1); Page 31-34; 60pp; Japanese.
XX
CC   This invention describes novel DNA sequences (I) encoding the human p51
CC   promoter domain (which may include the 5'-untranslated sequence) or
CC   derived from it by addition, deletion and/or substitution of one or more
CC   bases. The invention also describes (1) expression plasmids including (I)
CC   ; (2) host cells transformed by (1); (3) DNA probes binding to all or
CC   pert of (I); (4) cloning (I) using (3); (5) DNA (and RNA corresponding to
CC   it) inhibiting the expression of p51, which includes the antisense
```

```
CC  sequence of (I); (6) screening compounds for their ability to modify p51
CC  promoter activity, by observing their effect on cells transformed by (I);
CC  (7) compounds identified by (6); and (8) drug compositions containing
CC  (7). The products of the invention have cytostatic activity and can be
CC  used for gene therapy. (I) is used to screen for compounds that modify
CC  p51 promoter activity. (I) and RNA corresponding to it are used to
CC  inhibit the expression of p51. (I) and the identified compounds are used
CC  for the diagnosis and treatment of cancer
XX
SQ  Sequence 5676 BP; 1784 A; 1053 C; 1116 G; 1723 T; 0 U; 0 Other;

    Query Match        100.0%;  Score 5676;  DB 5;  Length 5676;
    Best Local Similarity  100.0%;  Pred. No. 0;
    Matches 5676;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy     1 CAGCTGTTCAGGGATGTCTGAAAAGAAGCCACCCACATTGCTCTGGACACTGGGTGT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAGCTGTTCAGGGATGTCTGAAAAGAAGCCACCCACATTGCTCTGGACACTGGGTGT 60

Qy    61 GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120

Qy   121 TGGCCTCGTCTTCCTTGCATGCCCTCCTTTCCTTGGGACACTCCTTAAGGCATCTTCTTG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGGCCTCGTCTTCCTTGCATGCCCTCCTTTCCTTGGGACACTCCTTAAGGCATCTTCTTG 180

Qy   181 ACATTAACTTAACTATAAATGTTTATTTGATGAATTCAGTGACCTGAAGAGAGATGGAG 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACATTAACTTAACTATAAATGTTTATTTGATGAATTCAGTGACCTGAAGAGAGATGGAG 240

Qy   241 GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA 300

Qy   301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGTTACTTACTCCTAATCGAT 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGTTACTTACTCCTAATCGAT 360

Qy   361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCCATGA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCCATGA 420

Qy   421 CAGAACTTTGGCTTCCACTTTACTCACACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTGTATGTA 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CAGAACTTTGGCTTCCACTTTACTCACACTTTACTAAAATAGAGATTGTGTGGTTGAGCTGCAGCTGTATGTA 480

Qy   481 CAGAAAAGTGTCATACATTAAAAATACCAGTCTCTTCAATTTGAGCAATAGT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAGAAAAGTGTCATACATTAAAAATACCAGTCTCTTCAATTTGAGCAATAGT 540

Qy   541 TGGTGAATTACTCACCACCTCCTCGAAGGTTCTTCTCCTGCTCCTCCACTATA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TGGTGAATTACTCACCACCTCCTCGAAGGTTCTTCTCCTGCTCCTCCACTATA 600

Qy   601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAAGGAAT 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACACAAAGGAAT 660

Qy   661 TCAGTTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCAGTTTCTGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT 720

Qy   721 CTGCTTAAGCCTCTGTGTGGCATCTGTTTTCCACTGCCCTATTGCTCT 780
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CTGCTTAAGCCTCTGTGTGGCATCTGTTTTCCACTGCCCTATTGCTCT 780

Qy   781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCCTCTGCTCAGTCAGTCATGGTAGGG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCCTCTGCTCAGTCAGTCATGGTAGGG 840

Qy   841 CCATCAGTCCACATGATCAGGCCTCCTCCTGCCCAAACATGGCATCTTTCTTTTTGGGAATT 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CCATCAGTCCACATGATCAGGCCTCCTCCTGCCCAAACATGGCATCTTTCTTTTTGGGAATT 900
```

```
Qy   901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCTGGTTACCGCCTAC 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCTGGTTACCGCCTAC 960

Qy   961 AGTGGCTTTGCAAAGTGACTGTCCATTCCTAAGTCCCTGAATTATAATTTAT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AGTGGCTTTGCAAAGTGACTGTCCATTCCTAAGTCCCTGAATTATAATTTAT 1020

Qy  1021 CCTGGTTACAGCCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTTATAGTCTGT 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CCTGGTTACAGCCCCTTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTTATAGTCTGT 1080

Qy  1081 GAATTTCATATTTCTTTCATACATTTCATGTTTGTTTGTTTGTTTGTGTGTTTTT 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAATTTCATATTTCTTTCATACATTTCATGTTTGTTTGTTTGTTTGTGTGTTTTT 1140

Qy  1141 GGCTTTAGGTAGGCAGAATCAGTTTCTGTGTTTATACCCAAGGAATCCTGATTGATACA 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GGCTTTAGGTAGGCAGAATCAGTTTCTGTGTTTATACCCAAGGAATCCTGATTGATACA 1200

Qy  1201 TCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 TCCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 1260

Qy  1261 TCTGGGAGTAAGTAGTACCAGAGCTCGTACTAACCCCAGGTTAGCCTACCAACTGCTTTACAC 1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 TCTGGGAGTAAGTAGTACCAGAGCTCGTACTAACCCCAGGTTAGCCTACCAACTGCTTTACAC 1320

Qy  1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAAGAAGCTACTATTTTGC 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAAGAAGCTACTATTTTGC 1380

Qy  1381 CAAAGACCTTCAGGAGGACCAAGACAAGTTCTCGGGATATGTGATGATTGAACTCTTAAAA 1440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CAAAGACCTTCAGGAGGACCAAGACAAGTTCTCGGGATATGTGATGATTGAACTCTTAAAA 1440

Qy  1441 AGTTTGTTGGGACTTCTGCGGCCATAATTGTGTATCTAAGACCAGATTCATTCTTAATAGCT 1500
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 AGTTTGTTGGGACTTCTGCGGCCATAATTGTGTATCTAAGACCAGATTCATTCTTAATAGCT 1500

Qy  1501 AAACAAACAAACAAGATCCACAGGTTCAGCGTTATAATAAGAGTGAATTACTGATAC 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 AAACAAACAAACAAGATCCACAGGTTCAGCGTTATAATAAGAGTGAATTACTGATAC 1560

Qy  1561 AGTTGACAACATGAACAATGAATATATCTCAGAAAATGAGCAAAAAAATCCAGAC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 AGTTGACAACATGAACAATGAATATATCTCAGAAAATGAGCAAAAAAATCCAGAC 1620

Qy  1621 ACAGAAGAATACGTACCATATGCCTGCATTTATGTGATATTCAGCATTGTCCAA 1680
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 ACAGAAGAATACGTACCATATGCCTGCATTTATGTGATATTCAGCATTGTCCAA 1680

Qy  1681 CATAGTAGCCAGTAGCCTCAGTAGGCTTCAAATTGATTAAAATTAAGTAAGA 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 CATAGTAGCCAGTAGCCTCAGTAGGCTTCAAATTGATTAAAATTAAGTAAGA 1740

Qy  1741 ATAAAAATTAGCTTCGACAACATGGCTACCACATCAGAC 1800
         |||||||||||||||||||||||||||||||||||||||
Db  1741 ATAAAAATTAGCTTCGACAACATGGCTACCACATCAGAC 1800

Qy  1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG 1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG 1860

Qy  1861 AAAATATACACATAATCTATAAAAACAAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG 1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 AAAATATACACATAATCTATAAAAACAAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG 1920

Qy  1921 TGAGGGGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA 1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 TGAGGGGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGTAATAGGGTTGTTGGA 1980
```

Left column

```
Qy  1981  ATCTTGATTGCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  ATCTTGATTGCGATGAAGGCTACTCGGTGTCTAATGTGTCACCTCCTCAGACTGAACACT  2040

Qy  2041  TGGAATTGGCGAATTTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  TGGAATTGGCGAATTTCATTGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA  2100

Qy  2101  AGTGTTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA  2160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  AGTGTTTTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA  2160

Qy  2161  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTTGCTAGCTTCCTGAAAACAC  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTTGCTAGCTTCCTGAAAACAC  2220

Qy  2221  AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATG  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATG  2280

Qy  2281  AGTCTTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAA  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  AGTCTTATTCAAAAATAGCTAGACTCCAGGAAGAAAGTTATAGGTGACCTTACACAAAAA  2340

Qy  2341  GTGCAGATATAATTAATTATGTGACAGTAGGACAGTGGGGAAGGTTTCTTTTATGTGAAAAA  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  GTGCAGATATAATTAATTATGTGACAGTAGGACAGTGGGGAAGGTTTCTTTTATGTGAAAAA  2400

Qy  2401  GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGGGAGTTT  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGTGAGTTT  2460

Qy  2461  TAAAACATTCCTGGTTGGAGGGCAGAATATGATCCAAGGCACAAAGAGCAACCAGAAAATA  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  TAAAACATTCCTGGTTGGAGGGCAGAATATGATCCAAGGCACAAAGAGCAACCAGAAAATA  2520

Qy  2521  TGCAACCTAGAGGAAGTGCATGAGAGGGAGCAGTTGTAAAATAATTTTCATGAATGTAA  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TGCAACCTAGAGGAAGTGCATGAGAGGGAGCAGTTGTAAAATAATTTTCATGAATGTAA  2580

Qy  2581  GTGAGAAGAATTTGTATCATAGAGACACCTGAGTTTGGCGAGCGCATGTGTTCTTGGCTCCTA  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  GTGAGAAGAATTTGTATCATAGAGACACCTGAGTTTGGCGAGCGCATGTGTTCTTGGCTCCTA  2640

Qy  2641  GGAGTCAAGAAGAACAAAGTGTCCCTTTCCTACGTTATGCTCAGTGGTCCAAGTCCAA  2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  GGAGTCAAGAAGAACAAAGTGTCCCTTTCCTACGTTATGCTCAGTGGTCCAAGTCCAA  2700

Qy  2701  AACACCTTTCCTTCCTTAAGTACTTTCTTTCTTCTTCTCTCCCCATACAAATCTAAAGTCTTCACA  2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  AACACCTTTCCTTCCTTAAGTACTTTCTTTCTTCTTCTCTCCCCATACAAATCTAAAGTCTTCACA  2760

Qy  2761  AACATCATTTAAACAGCAGGCAATTTCTGCGCAATTGCTTTTTCCTTGGAAAGGTCAGAAAGTTTCCAC  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  AACATCATTTAAACAGCAGGCAATTTCTGCGCAATTGCTTTTTCCTTGGAAAGTCTTGGAAAGTTTCCAC  2820

Qy  2821  GTACGTTATTATTACAATTTCTGCCTAAAAGACTCTAAAGTTGGAAAGTTTCTAAAGTTTCCAC  2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  GTACGTTATTATTACAATTTCTGCCTAAAAGACTCTAAAGTTGGAAAGTTTCTAAAGTTTCCAC  2880

Qy  2881  CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAACCTTAGTCCCCTAAGAGAAT  2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAACCTTAGTCCCCTAAGAGAAT  2940

Qy  2941  AAGGATGAACTATAAAATATAGAAGTAATTATGTAATTATATAATATGTAATTATATGCCCACTTATT  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  AAGGATGAACTATAAAATATAGAAGTAATTATGTAATTATATAATATGTAATTATATGCCCACTTATT  3000

Qy  3001  TTTCACTTGATCGTGATGGTGCATGCTACTGGTGTTCGTGTTCTGTTGAATTCTAGAGAGTTTG  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  TTTCACTTGATCGTGATGGTGCATGCTACTGGTGTTCGTGTTCTGTTGAATTCTAGAGAGTTTG  3060
```

Right column

```
Qy  3061  CCTCTTTTTCCTGGTCAACTCTGCCATTATTCCATAATGCAATAATAGAGAGCCAATCTT  3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  CCTCTTTTTCCTGGTCAACTCTGCCATTATTCCATAATGCAATAATAGAGAGCCAATCTT  3120

Qy  3121  TTTCATAATTACTTATTTAAAATTGTTGCCATTTAAATTTCTGTTCCTCTTAGCTTAGTA  3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  TTTCATAATTACTTATTTAAAATTGTTGCCATTTAAATTTCTGTTCCTCTTAGCTTAGTA  3180

Qy  3181  ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTAAATGATATTATT  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  ACTTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTAAATGATATTATT  3240

Qy  3241  TAAATACGTTAGGCTATAAACCTTTAAAAAATTTTTAAAAAATAGAGTGTGGTGGTGGCT  3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TAAATACGTTAGGCTATAAACCTTTAAAAAATTTTTAAAAAATAGAGTGTGGTGGTGGCT  3300

Qy  3301  CATGCCTGTAATCCCAACACTTTGGGAAGCCGGGGTCGGGGAGGATAGCTTGAGTCCAGCAG  3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  CATGCCTGTAATCCCAACACTTTGGGAAGCCGGGGTCGGGGAGGATAGCTTGAGTCCAGCAG  3360

Qy  3361  TTTGAGACCAGTCAGGGCAACACAGCAGAGACCCCATATCTAAAAAAACAAAACAAAAACAA  3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  TTTGAGACCAGTCAGGGCAACACAGCAGAGACCCCATATCTAAAAAAACAAAACAAAAACAA  3420

Qy  3421  AATTACCTGGGTATGGTGTGTCTCCACCTGAGTCCAAGCTACACAGGAAGCTGAGGCAGA  3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  AATTACCTGGGTATGGTGTGTCTCCACCTGAGTCCAAGCTACACAGGAAGCTGAGGCAGA  3480

Qy  3481  AGGATCACTTGAGCCCAGGAGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACTC  3540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  AGGATCACTTGAGCCCAGGAGTTGAGGCTGCAGTGATCCATGAACCGCGCTGCTACACTC  3540

Qy  3541  AGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAATAAATAATAACTTTT  3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3541  AGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAATAAATAATAACTTTT  3600

Qy  3601  AAAAAACAAAAATTAATTAATTAATTTTAAAAAACAACACACTAGAGATGTTTGCAAATTGA  3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  AAAAAACAAAAATTAATTAATTAATTTTAAAAAACAACACACTAGAGATGTTTGCAAATTGA  3660

Qy  3661  TTATTTGGGAGTCTATATCCCTGGAATTAATTAAAATATTAGAAGAGTTCTTCCTCA  3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  TTATTTGGGAGTCTATATCCCTGGAATTAATTAAAATATTAGAAGAGTTCTTCCTCA  3720

Qy  3721  TTTCCTAGAGACGTGCGAATTGTAAATATCAGAGCATCAAGTAACCTGGAAGCGTGTTAGAAATGCAG  3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  TTTCCTAGAGACGTGCGAATTGTAAATATCAGAGCATCAAGTAACCTGGAAGCGTGTTAGAAATGCAG  3780

Qy  3781  CCAAAGTGTGGTCCAAGGTCCAAGTAACCTGGAAGCGTGTTAGAAATGCAG  3840
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3781  CCAAAGTGTGGTCCAAGGTCCAAGTAACCTGGAAGCGTGTTAGAAATGCAG  3840

Qy  3841  AGTCTTAGGCCTCACCCAGATCTGCATTAACCAGAATCTGCACTAGAAATCTTCACTCCACCTTTC  3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3841  AGTCTTAGGCCTCACCCAGATCTGCATTAACCAGAATCTGCACTAGAAATCTTCACTCCACCTTTC  3900

Qy  3901  GCCTCACGGGCACATTAAAACTGGAGGCAGGGAAGAAAGCCAAACTCTTTTACT  3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3901  GCCTCACGGGCACATTAAAACTGGAGGCAGGGAAGAAAGCCAAACTCTTTTACT  3960

Qy  3961  ATTATAAATGGAATCACTTGGGCTGTGTCACAGGAATTGATTATTTTAATTTCAGAA  4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3961  ATTATAAATGGAATCACTTGGGCTGTGTCACAGGAAATTGATTATTTTAATTTCAGAA  4020

Qy  4021  CCTTCATTTAGGTCATCTATATTTGCTAATAGCAGGAAGCCAAACTCTTTTACT  4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGAAGAAGCCAAACTCTTTTACT  4080

Qy  4081  GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG  4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4081  GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG  4140

Qy  4141  TGGAGCAAGCTGTTTGATTTGGCTGGGCGCCGCCTGTTGTGAATTCACAATT  4200
```

Left column (positions 4141–5280):

```
Db  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    TGGAGCAAGCTGTTTGATTTGGCTGGGCCTCAGGCCGGCCGGGCCTGTTTGTGAATTTCACAATT  4200
Qy  4201 CACAGATGGTAGCCGCCTCTCGGGCTAAGTAAAGGAAGAGAGAATGTCAAGTTTAAATAGCT  4260
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4201 CACAGATGGTAGCCGCCTCTCGGGCTAAGTAAAGGAAGAGAGAATGTCAAGTTTAAATAGCT  4260
Qy  4261 TCTCCCTTCCATCCTGGCTGAAGCAACAACAACAATAATATTTTATGAAACACATTTTGAGT  4320
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4261 TCTCCCTTCCATCCTGGCTGAAGCAACAACAACAATAATATTTTATGAAACACATTTTGAGT  4320
Qy  4321 TAGATTTACTTACACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTTGTCTTCACAAC  4380
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4321 TAGATTTACTTACACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTTGTCTTCACAAC  4380
Qy  4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGGA  4440
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGGA  4440
Qy  4441 AGATGTGAACTCACCATGTTAGTACAGTTAGATACACAGAGTGGTTTTTTTTCCCCCTG  4500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441 AGATGTGAACTCACCATGTTAGTACAGTTAGATACACAGAGTGGTTTTTTTTCCCCCTG  4500
Qy  4501 TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATCCACAAA  4560
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501 TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATCCACAAA  4560
Qy  4561 ACCAGGATAAGTTTACAGCCCATATTGGGAAATAAAATATTTGTGTGTAGACTT  4620
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561 ACCAGGATAAGTTTACAGCCCATATTGGGAAATAAAATATTTGTGTGTAGACTT  4620
```

RESULT 2

```
AAF25907/c
ID   AAF25907 standard; RNA; 5676 BP.
XX
AC   AAF25907;
XX
DT   19-APR-2001 (first entry)
XX
DE   Human tumor suppressor gene p51 promoter associated RNA SEQ ID 4.
XX
KW   Tumor suppressor; p51; cell death; cell proliferation; cancer;
KW   cytostatic; gene therapy; screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100818-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-JP004261.
XX
PR   29-JUN-1999; 99JP-00183195.
XX
PA   (NIPK ) NIPPON KAYAKU KK.
PA   (SAKA/) SAKAI T.
XX
PI   Sakai T, Kagaya S, Sato T, Sukenaga Y, Fujii H;
XX
DR   WPI; 2001-112452/12.
XX
PT   New DNA sequences encoding the human p51 promoter domain for screening
PT   for compounds that modify p51 promoter activity and for diagnosing and
PT   treating cancer.
XX
PS   Disclosure; Page 42-45; 60pp; Japanese.
XX
CC   This invention describes novel DNA sequences (I) encoding the human p51
CC   promoter domain (which may include the 5'-untranslated sequence) or
CC   derived from it by addition, deletion and/or substitution of one or more
CC   bases. The invention also describes (1) expression plasmids including (I)
CC   ; (2) host cells transformed by (1);(3) DNA probes binding to all or
CC   pert of (I); (4) cloning (I) using (3); (5) DNA (and RNA corresponding to
```

Right column (positions 5221–5676):

```
Db  5221 TACAAGGAAGTGTTATGTAGCTGGAGCAGGGTGGACACTCATCAGCTCAGTTCAGTTACA  5280
Qy  5281 AAAGTCCAGGCTGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCCAATCACGACA  5340
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281 AAAGTCCAGGCTGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCCAATCACGACA  5340
Qy  5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGGCTACTA  5400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGGCTACTA  5400
Qy  5401 TACGTCAAGGACTTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA  5460
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401 TACGTCAAGGACTTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA  5460
Qy  5461 GCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA  5520
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461 GCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA  5520
Qy  5521 TCAAGAAACGCTCCGCTCTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTA  5580
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521 TCAAGAAACGCTCCGCTCTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTA  5580
Qy  5581 TGTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATACAC  5640
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581 TGTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCCGGCTTTATATCTATATACAC  5640
Qy  5641 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT  5676
    ||||||||||||||||||||||||||||||||||||
Db  5641 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT  5676
```

```
CC   it) inhibiting the expression of p51, which includes the antisense
CC   sequence of (I); (6) screening compounds for their ability to modify p51
CC   promoter activity, by observing their effect on cells transformed by (I);
CC   (7) compounds identified by (6); and (8) drug compositions containing
CC   (7). The products of the invention have cytostatic activity and can be
CC   used for gene therapy. (I) and RNA corresponding to it are used to
CC   inhibit the expression of p51. (I) and the identified compounds are used
CC   for the diagnosis and treatment of cancer
XX
SQ   Sequence 5676 BP; 1723 A; 1116 C; 1053 G; 0 T; 1784 U; 0 Other;

Query Match             100.0%;  Score 5676;  DB 5;  Length 5676;
Best Local Similarity   100.0%;
Matches 5676;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy     1 CAGCTGTTCAGGGATGTCTGAAAAGAAGCCACCCACCCACCATTGCTTCTGGACACTGGGTGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5676 CAGCTGTTCAGGGATGTCTGAAAAGAAGCCACCCACCCACCATTGCTTCTGGACACTGGGTGT 5617

Qy    61 GACTTTGGAGGGTATCAGGTTTGTCGTTAAAGAAACTGCCAACCTCTTCCTGCGCCCCAAAT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5616 GACTTTGGAGGGTATCAGGTTTGTCGTTAAAGAAACTGCCAACCTCTTCCTGCGCCCCAAAT 5557

Qy   121 TGGCCCTCGTGTCCCTTCCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5556 TGGCCCTCGTGTCCCTTCCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 5497

Qy   181 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5496 ACATTAACTTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 5437

Qy   241 GTCAAATCAGAAGAGCACATGGCTAAGGTTGCAAGTTCCAATGCAAGTTGATTGAATTAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5436 GTCAAATCAGAAGAGCACATGGCTAAGGTTGCAAGTTCCAATGCAAGTTGATTGAATTAA 5377

Qy   301 AGTCATTGAATACCATTCAGTTACTTAAGTTCTAGGCCCAGCTTTACTCCTAATCGAT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5376 AGTCATTGAATACCATTCAGTTACTTAAGTTCTAGGCCCAGCTTTACTCCTAATCGAT 5317

Qy   361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5316 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA 5257

Qy   421 CAGAACTTGGCTTCCACTTTACTAACTTTACTAACTTACCATTAAAATAGAGATTGTGTGGTTGAGCTGCGAGCTATGTA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5256 CAGAACTTGGCTTCCACTTTACTAACTTTACTAACTTACCATTAAAATAGAGATTGTGTGGTTGAGCTGCGAGCTATGTA 5197

Qy   481 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTTCTTCAATTTGAGCAATAGT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5196 CAGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTTCTTCAATTTGAGCAATAGT 5137

Qy   541 TGGTGAATTTACTCCACCACCTCCTCCTTGAAGGTTCTTTCCTGCTCCTGCTCCTGCTCCTCCACTATA 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5136 TGGTGAATTTACTCCACCACCTCCTCCTTGAAGGTTCTTTCCTGCTCCTGCTCCTGCTCCTCCACTATA 5077

Qy   601 AATGCAGCAGGATGACCTGGAAAGGCTAGGACCTCAGGTTCAGTTACCCTGACACAAAGGAAT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5076 AATGCAGCAGGATGACCTGGAAAGGCTAGGACCTCAGGTTCAGTTACCCTGACACAAAGGAAT 5017

Qy   661 TCAGTTTCTGATCTCATAGTCAGAGCTGCCAGAGCTCTACGGCTCTACGGAACATGCAAGATCAT 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5016 TCAGTTTCTGATCTCATAGTCAGAGCTGCCAGAGCTCTACGGCTCTACGGAACATGCAAGATCAT 4957

Qy   721 CTGCTTTAAGCCTCTGTGGTGCATCTGTTTTCCACTGCCCTGTGTTTTTCCACTGCCCTGTACCTATTGCTCT 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4956 CTGCTTTAAGCCTCTGTGGTGCATCTGTTTTCCACTGCCCTGTGTTTTTCCACTGCCCTGTACCTATTGCTCT 4897

Qy   781 TTCCTTGGTTAACAGAACCTTTATTTCTTCTGAAACTCTCTGCTCACTCATGCATGGTAGGG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4896 TTCCTTGGTTAACAGAACCTTTATTTCTTCTGAAACTCTCTGCTCACTCATGCATGGTAGGG 4837

Qy   841 CCATCAGTCCACATGATCAGCCGCCTCCTCCTGGCCAAACATGGCATCTTTCTTTTGGGAATT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4836 CCATCAGTCCACATGATCAGCCGCCTCCTCCTGGCCAAACATGGCATCTTTCTTTTGGGAATT 4777

Qy   901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAGAAAAAGCTGTTGAATCTGACTTACGCCTAC 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4776 TGAATCTTAAGCTGAATAGCTGAAGTTCAAGAAAAAGCTGTTGAATCTGACTTACGCCTAC 4717

Qy   961 AGTGGCTTTGCAAAGTGACTGTCCATTCCATTCCTTAAGCTCCCTGAATTTATAATTTAT 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4716 AGTGGCTTTGCAAAGTGACTGTCCATTCCATTCCTTAAGCTCCCTGAATTTATAATTTAT 4657

Qy  1021 CCTGGTTACAGCCCTTTCTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTATAGTCTGT 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4656 CCTGGTTACAGCCCTTTCTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTATAGTCTGT 4597

Qy  1081 GAATTTCATATTTCTTTCATACATTTTCATACATTTTCATGTTTTGTTTGTTTGTGTGTTTTTT 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4596 GAATTTCATATTTCTTTCATACATTTTCATACATTTTCATGTTTTGTTTGTTTGTGTGTTTTTT 4537

Qy  1141 GGCTTTAGGTAGGCAGAATCAGTTGTTCGTGTGTTTATACCAAGGAATCCTCAAGAGAGTAGGCTACCTGCCTGAGG 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4536 GGCTTTAGGTAGGCAGAATCAGTTGTTCGTGTGTTTATACCAAGGAATCCTCAAGAGAGTAGGCTACCTGCCTGAGG 4477

Qy  1201 TCCTTCCCTTAAAATAAAGTATCTAAGGTCAAAGAGCTCAAAGAGCTCTTCCCTTAAGGCATCTTCTTG 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4476 TCCTTCCCTTAAAATAAAGTATCTAAGGTCAAAGAGCTCAAAGAGCTCTTCCCTTAAGGCATCTTCTTG 4417

Qy  1261 TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAACCCCAGGTTAGCCAACTGCTTACAC 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4416 TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAACCCCAGGTTAGCCAACTGCTTACAC 4357

Qy  1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAGAAGCTACTATTTTGC 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4356 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAGAAGCTACTATTTTGC 4297

Qy  1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCGGGGATATGTGATGATATGTGATGATCCTTAAAA 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4296 CAAAGACCTCAGGAGGACCAAGAACAAGTTCGGGGATATGTGATGATATGTGATGATCCTTAAAA 4237

Qy  1441 AGTTTGTTGGACTTCTTCGGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4236 AGTTTGTTGGACTTCTTCGGCCATAATTGTGTATCTAAGACCAGATTTCATTCTTAATAGCT 4177

Qy  1501 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4176 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC 4117

Qy  1561 AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4116 AGTTGACAACATGAATATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC 4057

Qy  1621 ACAGAGAATACGTACCATATGCCTGCATTATGTGATTCTAGCATTGTATTGTCCAA 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4056 ACAGAGAATACGTACCATATGCCTGCATTATGTGATTCTAGCATTGTATTGTCCAA 3997

Qy  1681 CATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTGATTAAAATAAGTAAGA 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3996 CATAGTAGCCAGTAGCCTCACATGGCTATTCAAATTGATTAAAATAAGTAAGA 3937

Qy  1741 ATAAAAATTTAGCTCTTCAGTAGCGTTAGCGTTACCACCATCAGAC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  3936 ATAAAAATTTAGCTCTTCAGTAGCGTTAGCGTTACCACCATCAGAC 3877

Qy  1801 GGTGCAAATATGAATATTTCCTTTATAACAAAAGCAAGTTCATTGGAAAACAATGTTCTAG 1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3876 GGTGCAAATATGAATATTTCCTTTATAACAAAAGCAAGTTCATTGGAAAACAATGTTCTAG 3817

Qy  1861 AAAATATCACACATAATCTATAAAACAAAAACAAGCAAGTGCTCAGTAGTTGTCTAAGGCCAGGGG 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3816 AAAATATCACACATAATCTATAAAACAAAAACAAGCAAGTGCTCAGTAGTTGTCTAAGGCCAGGGG 3757

Qy  1921 TGAGGGGATCGATTGCAAAGTGGTATGAGGGAAAGTTTGGGTTGTTGTTGA 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3757 TGAGGGGATCGATTGCAAAGTGGTATGAGGGAAAGTTTGGGTTGTTGTTGA 1921
```

```
Db  3756 TGAGGGGAGATCGATTGCAAAGTGGTATGAGGAAAGTTTTGGGGTAATAGGGTTGTTGGA 3697

Qy  1981 ATCTTGATTGCGATGAAGGCTACTCGGTCGTCTAAATGTCTCAGACTGAACACT 2040
              ||||||||||||||||||||||||||||||||||||||||||||
Db  3696 ATCTTGATTGCGATGAAGGCTACTCGGTCGTCTAAATGTCTCAGACTGAACACT 3637

Qy  2041 TGGAATTGGGCGAATTTCATTCGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3636 TGGAATTGGGCGAATTTCATTCGTATGTAAATTATACCTCATAAAGTAACTCTAAGAGGTCA 3577

Qy  2101 AGTGTTTTGTTGGAAATTATTTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3576 AGTGTTTGTTGGAAATTATTTTTAATCAGTTGCAATACTTATTATGAGATGATTTTTGCA 3517

Qy  2161 AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTTCCTGAAAAACAC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3516 AATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTTCCTGAAAAACAC 3457

Qy  2221 AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATG 2280
              |||||||||||||||||||||||||||||||||||||||||||||||
Db  3456 AGAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATG 3397

Qy  2281 AGTCTATTCAAAAATAGCTAGACTCCAGAAGAAAGTTATAGGTGACCTTACACACAAAAA 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3396 AGTCTATTCAAAAATAGCTAGACTCCAGAAGAAAGTTATAGGTGACCTTACACACAAAAA 3337

Qy  2341 GTGCAGATAATAATTATGTAGGACAGTAGCAGTGGGGAAGTTTCTTTATGTGAATAAAA 2400
              |||||||||||||||||||||||||||||||||||||||||||||||
Db  3336 GTGCAGATATAATTATGTAGGACAGTAGCAGTGGGGAAGTTTCTTTATGTGAATAAAA 3277

Qy  2401 GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGAGTTT 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db  3276 GAGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGCAGATGAGTTT 3217

Qy  2461 TAAAACATTCCTGGTGGAGGCAGAAATGTCCCCTTTCTCCTCATGGTCTCCTACGGTCCAA 2520
              ||||||||||||||||||||||||||||||||||||||||||||||
Db  3216 TAAAACATTCCTGGTGGAGGCAGAAATGTCCCCTTTCTCCTCATGGTCTCCTACGGTCCAA 3157

Qy  2521 TGCAACCTAGAGGAAAGTGCATGAGGGGAGCAGTTGTAAAATAATTTTCATGGAATGTAA 2580
              ||||||||||||||||||||||||||||||||||||||||||||||
Db  3156 TGCAACCTAGAGGAAAGTGCATGAGGGGAGCAGTTGTAAAATAATTTTCATGGAATGTAA 3097

Qy  2581 GTGAGAAGAATTTGTATCATAGACACCTGAGTTTGCAGAGTGCATGTTCTTGGCTCCTA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||
Db  3096 GTGAGAAGAATTTGTATCATAGACACCTGAGTTTGCAGAGTGCATGTTCTTGGCTCCTA 3037

Qy  2641 GGAGTCAAGAAGAACAAAGTGTCCCCTTTCTCCTACGGTCATGCTCCAGTGGTCCAAGTCCAA 2700
              |||||||||||||||||||||||||||||||||||||||||||||
Db  3036 GGAGTCAAGAAGAACAAAGTGTCCCCTTTCTCCTACGGTCATGCTCCAGTGGTCCAAGTCCAA 2977

Qy  2701 AACACCTTTCCTTAAGTACTTCTTCTTCCCCATCACAAATCTAAAGTCTTCACA 2760
              |||||||||||||||||||||||||||||||||||||||||||
Db  2976 AACACCTTTCCTTAAGTACTTCTTCTTCCCCATCACAAATCTAAAGTCTTCACA 2917

Qy  2761 AACATCATTTAAACAGGCAGTCATGGTCAGAAAGCAATTGCTTTTCCTTTTCCTAGACTTCTAT 2820
              |||||||||||||||||||||||||||||||||||||||||||||
Db  2916 AACATCATTTAAACAGGCAGTCATGGTCAGAAAGCAATTGCTTTTCCTTTTCCTAGACTTCTAT 2857

Qy  2821 GTACGTTAATTATATACAATTTCTGCCTAAAGACTCTAAAGACTCTAAAGTCTTCCCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2856 GTACGTTAATTATATACAATTTCTGCCTAAAGACTCTAAAGACTCTAAAGTCTTCCCAC 2797

Qy  2881 CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAGTAACCTTGTATAGTCCCCTAAGAGAAT 2940
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2796 CTTGCACATCAAAGATATAATTCATGCATTTGTATAGTAGTAACCTTGTATAGTCCCCTAAGAGAAT 2737

Qy  2941 AAGGATGAACTACGTTATTATACAATTTGTAATTATAATATAAGAAGTAATATGGTAATTAAT 3000
              |||||||||||||||||||||||||||||||||||||||||||
Db  2736 AAGGATGAACTACGTTATTATACAATTTGTAATTATAATATAAGAAGTAATATGGTAATTAAT 2677

Qy  3001 TTTCACTTGATCGTCGTATGGTTGCATGCGTTGCATGCTACTGGGTTCTGGTTGTTCTGTTGGA 3060
              |||||||||||||||||||||||||||||||||||||||||||||
Db  2676 TTTCACTTGATCGTCGTATGGTTGCATGCGTTGCATGCTACTGGGTTCTGGTTGTTCTGTTGGA 2617
```

```
Qy  3061 CCTCTTTTTCCTGGTCAACTCTCGCCATTTATTTCCATAATGCAATATAGAGCCAATCTT 3120
              |||||||||||||||||||||||||||||||||||||||||||||||
Db  2616 CCTCTTTTTCCTGGGTCAACTCTCGCCATTTATTTCCATAATGCAATATAGAGCCAATCTT 2557

Qy  3121 TTTCATAATTACTTATTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA 3180
              ||||||||||||||||||||||||||||||||||||||||||||||
Db  2556 TTTCATAATTACTTATTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTA 2497

Qy  3181 ACTTTAGGATTTTAAATAACAACTATTGAAATCATTGAAATCATGACATACGTTAAATGATATTATT 3240
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2496 ACTTTAGGATTTTAAATAACAACTATTGAAATCATTGAAATCATGACATACGTTAAATGATATTATT 2437

Qy  3241 TAAATACGTTAGGCTATAAACCTTTTAAATTTTTAAAAAATTTTAAAAAAATAGAGTGTGGTGGCT 3300
              |||||||||||||||||||||||||||||||||||||||||||||
Db  2436 TAAATACGTTAGGCTATAAACCTTTTAAATTTTTAAAAAATTTTAAAAAAATAGAGTGTGGTGGCT 2377

Qy  3301 CATGCCTGTAATCCAACACTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAG 3360
              |||||||||||||||||||||||||||||||||||||||||||
Db  2376 CATGCCTGTAATCCAACACTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAG 2317

Qy  3361 TTTGAGACCAGTCAGGCAACACAGCAAGAAGACCCCATATCTCAAAAAAACAAAACAAACAA 3420
              |||||||||||||||||||||||||||||||||||||||||||
Db  2316 TTTGAGACCAGTCAGGCAACACAGCAAGAAGACCCCATATCTCAAAAAAACAAAACAAAACAA 2257

Qy  3421 AATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAGCTGAGGCAGA 3480
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2256 AATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAGCTGAGGCAGA 2197

Qy  3481 AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGACCGCGCTGCTACACTC 3540
              |||||||||||||||||||||||||||||||||||||||||||
Db  2196 AGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGACCGCGCTGCTACACTC 2137

Qy  3541 AGTCTGGTGACAGTGCAGTGCTCTCAAAAATAATAATAAAAATAACTTTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2136 AGTCTGGTGACAGTGCAGTGCTCTCAAAAATAATAATAAAAATAACTTTT 2077

Qy  3601 AAAAAACAAAAATTAATTAAATTATTAATTTAAAAAACACAACACACTAGACATGTTTGCAAATTGA 3660
              |||||||||||||||||||||||||||||||||||||||||||||
Db  2076 AAAAAACAAAAATTAATTAAATTATTAATTTAAAAAACACAACACACTAGACATGTTTGCAAATTGA 2017

Qy  3661 TTATTTGGGAGTCTATATCCCTGAAGTTAATTTAAAAATATTTAGAGAGTTCTTCCTCA 3720
              ||||||||||||||||||||||||||||||||||||||||||||
Db  2016 TTATTTGGGAGTCTATATCCCTGAAGTTAATTTAAAAATATTTAGAGAGTTCTTCCTCA 1957

Qy  3721 TTTCCTAGAGACGTCGAATTGTAAATATCAGAGCTTAGAAGGAACACTAGGGCTCGCCACT 3780
              |||||||||||||||||||||||||||||||||||||||||||||
Db  1956 TTTCCTAGAGACGTCGAATTGTAAATATCAGAGCTTAGAAGGAACACTAGGGCTCGCCACT 1897

Qy  3781 CCAAAGTGTGGTCCAAGGACCAGCAGCATCAAGTAACCTGGGACGTGTTAGAAATGCAG 3840
              ||||||||||||||||||||||||||||||||||||||||||||
Db  1896 CCAAAGTGTGGTCCAAGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAG 1837

Qy  3841 AGTCTTAGGCCTCACCCCAGAATCTGCATTAACAGAATTCTTAGGT 3900
              ||||||||||||||||||||||||||||||||||||||||||
Db  1836 AGTCTTAGGCCTCACCCCAGAATCTGCATTAACAGAATTCTTAGGT 1777

Qy  3901 GCCTCACGGGCACATTAAAACTTAAAGACTTGAGAAATCTTCACTCCACCTTTC 3960
              ||||||||||||||||||||||||||||||||||||||||||
Db  1776 GCCTCACGGGCACATTAAAACTTAAAGACTTGAGAAATCTTCACTCCACCTTTC 1717

Qy  3961 ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATGATTATTTTTAATTTCAGAA 4020
              |||||||||||||||||||||||||||||||||||||||||||
Db  1716 ATTATAAATGGAATCACTTGGGCTGTGGTCACAGGAAATGATTATTTTTAATTTTCAGAA 1657

Qy  4021 CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGAAGAAAGCCAAACTCTTTAACT 4080
              ||||||||||||||||||||||||||||||||||||||||||
Db  1656 CCTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGAAGAAAGCCAAACTCTTTAACT 1597

Qy  4081 GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTACATTTTG 4140
              ||||||||||||||||||||||||||||||||||||||||||
Db  1596 GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTACATTTTG 1537
```

```
Qy  4141 TGGAGCAAGCTGTTTGATTTGGCTCGGGGCTCAGGCCCGGCCTCGTTTGTGAATTCACAATT 4200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1536 TGGAGCAAGCTGTTTGATTTGGCTCGGGGCTCAGGCCCGGCCTCGTTTGTGAATTCACAATT 1477

Qy  4201 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGAAGAGAAATGTCAAGTTTAAATAGCT 4260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1476 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGAAGAGAAATGTCAAGTTTAAATAGCT 1417

Qy  4261 TCTCCCCTTCCATCCTGGCTGGAAGCAACAATAAAATATTTTATGAAAACACATTTTGAGT 4320
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1416 TCTCCCTTCCATCCTGGCTGGAAGCAACAATAAAATATTTTATGAAAACACATTTTGAGT 1357

Qy  4321 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 4380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1356 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 1297

Qy  4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGTGAAGGGA 4440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1296 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGTGAAGGGA 1237

Qy  4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTCCCCCTG 4500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1236 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTCCCCCTG 1177

Qy  4501 TTGGGAGTCTATCCTAACTGAGCTTCTGAATCATATTCATTCAATTTCCAAATCCACAAA 4560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1176 TTGGGAGTCTATCCTAACTGAGCTTCTGAATCATATTCATTCAATTTCCAAATCCACAAA 1117

Qy  4561 ACCAGGATAAGTTTACAGCCCATATTCAGAAAGGAAATAAATTATTTGTGTGTAGACTT 4620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1116 ACCAGGATAAGTTTACAGCCCATATTCAGAAAGGAAATAAATTATTTGTGTGTAGACTT 1057

Qy  4621 TCCTGATATTACACTGATTTGGGAATATATGAACAATTTTATGTTTCCTTTCGAAGTAG 4680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1056 TCCTGATATTACACTGATTTGGGAATATATGAACAATTTTATGTTTCCTTTCGAAGTAG 997

Qy  4681 GTCAAGTCAAAGCAAAACCAAAACAGCAAAAACTGTAAGACATAAAGAGATAGAGTGGAG 4740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   996 GTCAAGTCAAAGCAAAACCAAAACAGCAAAAACTGTAAGACATAAAGAGATAGAGTGGAG 937

Qy  4741 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 4800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   936 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 877

Qy  4801 GTGCACTTCAGAATATTCTACAATAATAATTATTTCCAATTTTAATATCTTTAAGAAAA 4860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   876 GTGCACTTCAGAATATTCTACAATAATAATTATTTCCAATTTTAATATCTTTAAGAAAA 817

Qy  4861 TTATATATATGTAAGTACATGTGCATTGTTTGAGGTAGGATATTTAACTCAATAA 4920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   816 TTATATATATGTAAGTACATGTGCATTGTTTGAGGTAGGATATTTAACTCAATAA 757

Qy  4921 AGGTTATTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGATATCT 4980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   756 AGGTTATTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGATATCT 697

Qy  4981 CTTTTCTTAGCTGAGAGGAAGAGTGAGTTCCTAAGGTCTCAAGTTAAATAATCAAGGAAATTTCCCT 5040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   696 CTTTTCTTAGCTGAGAGGAAGAGTGAGTTCCTAAGGTCTCAAGTTAAATAATCAAGGAAATTTCCCT 637

Qy  5041 GTCTTTGCTATTTGAGATTGTGACCAACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 5100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   636 GTCTTTGCTATTTGAGATTGTGACCAACAGGCGGTTGGCTGAAAGGGAAACTGAAGGG 577

Qy  5101 CGGGGAGGAGGGGAAATAGATAGAAAAAACAAAACAAAACTTCCCTAAGCAGCTCT 5160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   576 CGGGGAGGAGGGGAAATAGATAGAAAAAACAAAACAAAACTTCCCTAAGCAGCTCT 517

Qy  5161 ACAAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCCAGA 5220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   516 ACAAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCCAGA 457

Qy  5221 TACAAGGAAGTGTTATGTAGCTGGAGCACTCATCAGCTCAGTTCAGTTACA 5280
```

```
Qy  5221 TACAAGGAAGTGTTATGTAGCTGGAGCACTCATCAGCTCAGTTCAGTTACA
Db   456 TACAAGGAAGTGTTATGTAGCTGGAGCACTCATCAGCTCAGTTCAGTTACA 397
         |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCATCCACGACA 5340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   396 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCATCCACGACA 337

Qy  5341 GAGATCAGAAGTTCAGAGAGTCCTCCAGCTCCAGTCCAAATTGCCAACAACAAGCAAGTGGCTACTA 5400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   336 GAGATCAGAAGTTCAGAGAGTCCTCCAGCTCCAGTCCAAATTGCCAACAACAAGCAAGTGGCTACTA 277

Qy  5401 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAGAAGAACAACAGTAGAGAGGATGCCCA 5460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   276 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAGAAGAACAACAGTAGAGAGGATGCCCA 217

Qy  5461 GCTGGTAAGAATCGAGTGTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAA 5520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   216 GCTGGTAAGAATCGAGTGTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAA 157

Qy  5521 TCAAGAAACCGTCCGCCTCTTTGCAAATATGTGCTTTTATATTGAAGGAGAAGTGCCTAAACTTCTA 5580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   156 TCAAGAAACCGTCCGCCTCTTTGCAAATATGTGCTTTTATATTGAAGGAGAAGTGCCTAAACTTCTA 97

Qy  5581 TGTCTGATAGCATTTGACCCTATTGTTTTAGCCTCCGGCTTTATATCTATATATACAC 5640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    96 TGTCTGATAGCATTTGACCCTATTGTTTTAGCCTCCGGCTTTATATCTATATATACAC 37

Qy  5641 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT 5676
         ||||||||||||||||||||||||||||||||||||
Db    36 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT 1
```

RESULT 3
AAF25906/c
ID   AAF25906 standard; DNA; 5676 BP.
XX
AC   AAF25906;
XX
DT   19-APR-2001 (first entry)
XX
DE   Human tumor suppressor gene p51 promoter associated DNA SEQ ID 3.
XX
KW   Tumor suppressor; p51; cell death; cell proliferation; cancer;
KW   cytostatic; gene therapy; screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100818-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-JP004261.
XX
PR   29-JUN-1999; 99JP-00183195.
XX
PA   (NIPK) NIPPON KAYAKU KK.
PA   (SAKA/) SAKAI T.
XX
PI   Sakai T, Kagaya S, Sato T, Sukenaga Y, Fujii H;
XX
DR   WPI; 2001-112452/12.
XX
PT   New DNA sequences encoding the human p51 promoter domain for screening
PT   for compounds that modify p51 promoter activity and for diagnosing and
PT   treating cancer.
XX
PS   Disclosure; Page 39-42; 60pp; Japanese.
XX
CC   This invention describes novel DNA sequences (I) encoding the human p51
CC   promoter domain (which may include the 5'-untranslated sequence) or
CC   derived from it by addition, deletion and/or substitution of one or more
CC   bases. The invention also describes (1) expression plasmids including (I)
CC   ; (2) host cells transformed by (1); (3) DNA probes binding to all or

```
CC  pert of (I); (4) cloning (I) using (3); (5) DNA (and RNA corresponding to
CC  it) inhibiting the expression of p51, which includes the antisense
CC  sequence of (I); (6) screening compounds for their ability to modify p51
CC  promoter activity, by observing their effect on cells transformed by (I);
CC  (7) compounds identified by (6); and (8) drug compositions containing
CC  (7). The products of the invention have cytostatic activity and can be
CC  used for gene therapy. (I) and RNA corresponding to it are used to
CC  inhibit the expression of p51. (I) and the identified compounds are used
CC  for the diagnosis and treatment of cancer
XX
SQ  Sequence 5676 BP; 1723 A; 1116 C; 1053 G; 1784 T; 0 U; 0 Other;

Query Match    100.0%;  Score 5676;  DB 5;  Length 5676;
Best Local Similarity    100.0%;
Matches 5676;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 CAGCTGTTCAGGGATGCTCTGGAAAGAAGCCACCACCACATTGCTCTCGACACTGGGTGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5676 CAGCTGTTCAGGGATGCTCTGGAAAGAAGCCACCACCACATTGCTCTCGACACTGGGTGT 5617

Qy     61 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCGCCCAAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5616 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCGCCCAAT 5557

Qy    121 TGGCCCTCGTGTTCCCTTGCCATGCCCTCTTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5556 TGGCCCTCGTGTTCCCTTGCCATGCCCTCTTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 5497

Qy    181 ACATTAACTTAACTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5496 ACATTAACTTAACTAACTATAAATGTTTATTTGATGAATTTCAGTGACCTGAAGAGATGGAG 5437

Qy    241 GTCAAATCAGAGAGCACATGGCTAAGGTGCAATGCACTTGCTTTTCATTGAATTAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5436 GTCAAATCAGAGAGCACATGGCTAAGGTGCAATGCACTTGCTTTTCATTGAATTAA 5377

Qy    301 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGCTTACTCCTAATCGAT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5376 AGTCATTCGAATACCATTCAGTTTACTTAAGTTCTAGGCCCAGCTTACTCCTAATCGAT 5317

Qy    361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5316 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGCAGGATCCTCTCCATGA 5257

Qy    421 CAGAACTTTGGCTTCCACTTTACTTACTTACTAAAATAAAAATCACCAAACTCAGTCTCAATTTGAGCAATAGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5256 CAGAACTTTGGCTTCCACTTTACTTACTTACTAAAATAAAAATCACCAAACTCAGTCTCAATTTGAGCAATAGT 5197

Qy    481 CAGAAAAGTGTCATACAATCAGTTCACCAAACTCAGTCTCTCAATTTGAGCAATAGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5196 CAGAAAAGTGTCATACAATCAGTTCACCAAACTCAGTCTCTCAATTTGAGCAATAGT 5137

Qy    541 TGGTGAATTTACTCCACCACCTCCTCCTTGAAGGTTCTTTCTGCTCCTGACCACAAAGGAAT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5136 TGGTGAATTTACTCCACCACCTCCTCCTTGAAGGTTCTTTCTGCTCCTGACCACAAAGGAAT 5077

Qy    601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACCAGTTACCACCATCAGATCAT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5076 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTACCCTGACCAGTTACCACCATCAGATCAT 5017

Qy    661 TCAGTTTCTCTGATCTCATAGTTCACAGGCTGCACAGGCTGCACCTGTTTGTTCCACTGCCCTGTACCTATTGCTCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5016 TCAGTTTCTCTGATCTCATAGTTCACAGGCTGCACAGGCTGCACCTGTTTGTTCCACTGCCCTGTACCTATTGCTCT 4957

Qy    721 CTGCTTTAAGCCTTCATAGTCCATGGATCGTGTGTGTCAGCATCTGTTCCACTGCCCTGTACCTATTGCTCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4956 CTGCTTTAAGCCTTCATAGTCCATGGATCGTGTGTGTCAGCATCTGTTCCACTGCCCTGTACCTATTGCTCT 4897

Qy    781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCTGCTCGTCAGTCATGTCATGGTAGGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4896 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTCTGCTCGTCAGTCATGTCATGGTAGGG 4837
```

```
Qy    841 CCATCAGTCCACATGATCAGGCCTCTTCCTCGGCCAAACATGGCATCTTCTTTGGGAATT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4836 CCATCAGTCCACATGATCAGGCCTCTTCCTGGCCAAACATGGCATCTTCTTTGGGAATT 4777

Qy    901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCTGAATCGACTTACGGCCTAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4776 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAGCTGTTGAATCTGAATCGACTTACGGCCTAC 4717

Qy    961 AGTGGCTTGCAAAGTGACTGCTCCATTCCTTAAGTCCCTGAATTTATAATTTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4716 AGTGGCTTGCAAAGTGACTGCTCCATTCCTTAAGTCCCTGAATTTATAATTTAT 4657

Qy   1021 CCTGGTTACAGCCCTTTCTGAGATGTGGTTTTTTTTCCAACTGTCTCTTATAGTCTGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4656 CCTGGTTACAGCCCTTTCTGAGATGTGGTTTTTTTTCCAACTGTCTCTTATAGTCTGT 4597

Qy   1081 GAATTTCATATTTCTTTCATACATTTCATGTTTGTTTGTTTGTTGTTGTGTGTTTTT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4596 GAATTTCATATTTCTTTCATACATTTCATGTTTGTTTGTTTGTTGTTGTGTGTTTTT 4537

Qy   1141 GGCTTTAGGTAGGCAGAATCAGTTTCTGTTTATACCCAAGGAATCCTGATTGATACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4536 GGCTTTAGGTAGGCAGAATCAGTTTCTGTTTATACCCAAGGAATCCTGATTGATACA 4477

Qy   1201 TCCTTCCCTTAAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4476 TCCTTCCCTTAAAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 4417

Qy   1261 TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAACCCCAGCTTAGCCAACTGCTTTACAC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4416 TCTGGGAGTAAGTTAGTACCAGAGCTCGTACTAACCCCAGCTTAGCCAACTGCTTTACAC 4357

Qy   1321 AACATTTGCTCTCTCCTTCAGAGTTCTTGGAGAAGAAGCTACTATTTTGC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4356 AACATTTGCTCTCTCCTTCAGAGTTCTTGGAGAAGAAGCTACTATTTTGC 4297

Qy   1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCGGGATATGTGATGATTGAACTCTAAAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4296 CAAAGACCTCAGGAGGACCAAGAACAAGTTCGGGATATGTGATGATTGAACTCTAAAA 4237

Qy   1441 AGTTTGTTGGACTTCTGGCCCATAATTGTGTATCTTAAGACCAGATTTCATTCTTAATAGCT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4236 AGTTTGTTGGACTTCTGGCCCATAATTGTGTATCTTAAGACCAGATTTCATTCTTAATAGCT 4177

Qy   1501 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATATAATAAGAGTGAATTACTGATAC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4176 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATATAATAAGAGTGAATTACTGATAC 4117

Qy   1561 AGTTGACAACATGATATCTCAGAAAATCCAGTCTTCAATTTGAGCAATAGT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4116 AGTTGACAACATGATATCTCAGAAAATCCAGTCTTCAATTTGAGCAATAGT 4057

Qy   1621 ACAGAAGAATACGTACCCATATGCCTGCATTTATGTGATATTCTAGCATTGATATTCTAGCATTGTCCAA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4056 ACAGAAGAATACGTACCCATATGCCTGCATTTATGTGATATTCTAGCATTGATATTCTAGCATTGTCCAA 3997

Qy   1681 CATAGTAGCCAGTTAGCTCTTCAGTTGATTAAGGCTATTCAAATTTGATTGATTAAATTAAGTAAGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3996 CATAGTAGCCAGTTAGCTCTTCAGTTGATTAAGGCTATTCAAATTTGATTGATTAAATTAAGTAAGA 3937

Qy   1741 ATAAAAATTTAGCTCTTCAGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTTCTAG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3936 ATAAAAATTTAGCTCTTCAGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTTCTAG 3877

Qy   1801 GGTGCAAATATGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTTCTAG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3876 GGTGCAAATATGAATATTTCCTTTATAACAGAAAGTTCATTGGAAAACAATGTTCTAG 3817

Qy   1861 AAAAATATACACATAATCTATAAAAACAAAAGCAGTCAGTCGATTGTCTAAGGCCAGGGGG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3816 AAAATATACACATAATCTATAAAAACAAAAGCAGTCAGTCGATTGTGATTGTCTAAGGCCAGGGGG 3757

Qy   1921 TGAGGGGGATCGATTGCAAAGTGGTATGAGGAGAAAGTTTTGGGGTTAATAGGGTTGTTGGA 1980
```

```
Qy  5221 TACAAGGAAGTGTTATGTAGCTGGAGCAGGTGGACACTCATCAGTCAGTTCAGTTACA 5280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   456 TACAAGGAAGTGTTATGTAGCTGGAGCAGGTGGACACTCATCAGTCAGTTCAGTTACA 397

Qy  5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCAATCCAGCATCAGCACGACA 5340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   396 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCAATCCAGCATCAGCACGACA 337

Qy  5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAACAACCAAGTGCTGGCTACTA 5400
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   336 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAACAACCAAGTGCTGGCTACTA 277

Qy  5401 TACGTCAAGGACTCTGAAGCCGTGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA 5460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   276 TACGTCAAGGACTCTGAAGCCGTGAGAGGGGAAGAACAACAGTAGAGAGGATGCCCA 217

Qy  5461 GCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA 5520
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   216 GCTGGTAAGAATCGAGTGTTTATGAAGTTTAGTCAATTGATGAATCTCATTGGCTAAAA 157

Qy  5521 TCAAGAAACGCTCCGCCTCTTGCAAATATGTAGGAGGAAGTGCTCAAACTTCTA 5580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   156 TCAAGAAACGCTCCGCCTCTTGCAAATATGTAGGAGGAAGTGCTCAAACTTCTA 97

Qy  5581 TGTCTGATAGCATTTGACCCTATTGCTTTAGCCTTTATATCTATATATACAC 5640
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    96 TGTCTGATAGCATTTGACCCTATTGCTTTAGCCTTTATATCTATATATACAC 37

Qy  5641 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT 5676
         ||||||||||||||||||||||||||||||||||||
Db    36 AGGTATTTGTGTATATTTTATATAATTGTTCTCCGT 1
```

RESULT 4
AAF25905
ID   AAF25905   standard; DNA; 5960 BP.
XX
AC   AAF25905;
XX
DT   19-APR-2001 (first entry)
XX
DE   Human tumor suppressor gene p51 promoter associated DNA SEQ ID 2.
XX
KW   Tumor suppressor; p51; cell death; cell proliferation; cancer;
KW   cytostatic; gene therapy; screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100818-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-JP004261.
XX
PR   29-JUN-1999; 99JP-00183195.
XX
PA   (NIPK) NIPPON KAYAKU KK.
PA   (SAKA/) SAKAI T.
XX
PI   Sakai T, Kagaya S, Sato T, Sukenaga Y, Fujii H;
XX
DR   WPI; 2001-112452/12.
XX
PT   New DNA sequences encoding the human p51 promoter domain for screening
PT   for compounds that modify p51 promoter activity and for diagnosing and
PT   treating cancer.
XX
PS   Claim 1(4); Page 35-38; 60pp; Japanese.
XX
CC   This invention describes novel DNA sequences (I) encoding the human p51
CC   promoter domain (which may include the 5'-untranslated sequence) or
CC   derived from it by addition, deletion and/or substitution of one or more
CC   bases. The invention also describes (I) expression plasmids including (I)

```
Qy  4141 TGGAGCAAGCTGTTGATTTGGCTGTGGGGCTCAGGCCGGCCGGCCCTGTTTGTGAATTTCACAATT 4200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1536 TGGAGCAAGCTGTTGATTTGGCTGTGGGGCTCAGGCCGGCCGGCCCTGTTTGTGAATTTCACAATT 1477

Qy  4201 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAAGAGAATGTCAAGTTTAAATAGCT 4260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1476 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAAGAGAATGTCAAGTTTAAATAGCT 1417

Qy  4261 TCTCCCTTCCATCCTGGCTGAAGCAACAACAAATAAATATTTTATGAAACACATTTTGAGT 4320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1416 TCTCCCTTCCATCCTGGCTGAAGCAACAACAAATAAATATTTTATGAAACACATTTTGAGT 1357

Qy  4321 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 4380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1356 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 1297

Qy  4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTCCTGTGTCCTGTGACTAGGGGTGAAGGGA 4440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1296 TTTGACATCTACTGATGTCACCTATTTACAGGTGTCCTGTGTCCTGTGACTAGGGGTGAAGGGA 1237

Qy  4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGTTTTTTTCCCCTG 4500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1236 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGTTTTTTTCCCCTG 1177

Qy  4501 TTGGGAGTCTATCCTCAACTGAGCTTCTGAATCATATTCATTCAATTCCAAATCCACAAA 4560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1176 TTGGGAGTCTATCCTCAACTGAGCTTCTGAATCATATTCATTCAATTCCAAATCCACAAA 1117

Qy  4561 ACCAGGATAAGTTTACAGCCCCATATTCAGGAATATATGAACAATTTATGGTTTCCTTTCGAAGTAG 4620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1116 ACCAGGATAAGTTTACAGCCCCATATTCAGGAATATATGAACAATTTATGGTTTCCTTTCGAAGTAG 1057

Qy  4621 TCCTGATATTACACTGATTTGGGAATATATGGAATAATGAACAATTTATGGTTTCCTTTCGAAGTAG 4680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1056 TCCTGATATTACACTGATTTGGGAATATATGGAATAATGAACAATTTATGGTTTCCTTTCGAAGTAG 997

Qy  4681 GTCAAGTCAAAGCAAAACAGCAAAAACCAAAAACAGCAAAAACTGTAAGACAATAAAAGATAGAGTGGAG 4740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   996 GTCAAGTCAAAGCAAAACAGCAAAAACCAAAAACAGCAAAAACTGTAAGACAATAAAAGATAGAGTGGAG 937

Qy  4741 CCGACTGAGAGAATTAAAATAAAACTAGAACTAGAATATTTTATTAACAGGCAATTTGAAATAATT 4800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   936 CCGACTGAGAGAATTAAAATAAAACTAGAACTAGAATATTTTATTAACAGGCAATTTGAAATAATT 877

Qy  4801 GTGCACTTCAGAATATTCTACAATATATATATATATATTTATATCTTTAAGAAAA 4860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   876 GTGCACTTCAGAATATTCTACAATATATATATATATATTTATATCTTTAAGAAAA 817

Qy  4861 TTACTATATATATGTAAGTACATGTGCATTGTGTTGAGGTAGCATGTGTTTGAGGTAGTAGATATTTAACTCAATAA 4920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   816 TTACTATATATATGTAAGTACATGTGCATTGTGTTGAGGTAGCATGTGTTTGAGGTAGTAGATATTTAACTCAATAA 757

Qy  4921 AGGTTATTTCTTTTTCTTTTATTCGGGTCAGGCAAGCTTCTAAGGGGATGTGAAAGGGATATCT 4980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   756 AGGTTATTTCTTTTTCTTTTATTCGGGTCAGGCAAGCTTCTAAGGGGATGTGAAAGGGATATCT 697

Qy  4981 CTTTCCTTAGCTGAGGAGGAGAAGAGTGAGTTCTAAGTAAATATAATCAAGGAATTTCCCT 5040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   696 CTTTCCTTAGCTGAGGAGGAGAAGAGTGAGTTCTAAGTAAATATAATCAAGGAATTTCCCT 637

Qy  5041 GTCTTTGCTATTTGAGATTGTGACCCACACAGCGCGGTTGGCTGAAGGGAAACTGAAGGG 5100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   636 GTCTTTGCTATTTGAGATTGTGACCCACACAGCGCGGTTGGCTGAAGGGAAACTGAAGGG 577

Qy  5101 CGGGGAGGAGGGGAAATAGATGAAAAACAAAACAAAACTTCCCTAAGCAGCTCT 5160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   576 CGGGGAGGAGGGGAAATAGATGAAAAACAAAACAAAACTTCCCTAAGCAGCTCT 517

Qy  5161 ACAAAACATTTTAGCCCCAGAAAATAGTCACAGAAAATCCTCAAATCAAACCAGTATCCAGA 5220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   516 ACAAAACATTTTAGCCCCAGAAATAGTCACAGAAAATCCTCAAATCAAACCAGTATCCAGA 457
```

```
CC   ; (2) host cells transformed by (1); (3) DNA probes binding to all or
CC   pert of (I); (4) cloning (I) using (3); (5) DNA (and RNA corresponding to
CC   it) inhibiting the expression of p51, which includes the antisense
CC   sequence of (I); (6) screening compounds for their ability to modify p51
CC   promoter activity, by observing their effect on cells transformed by (I);
CC   (7) compounds identified by (6); and (8) drug compositions containing
CC   (7). The products of the invention have cytostatic activity and can be
CC   used for gene therapy. (I) is used to screen for compounds that modify
CC   p51 promoter activity. (I) and RNA corresponding to it are used to
CC   inhibit the expression of p51. (I) and the identified compounds are used
XX   for the diagnosis and treatment of cancer
SQ   Sequence 5960 BP; 1858 A; 1103 C; 1175 G; 1824 T; 0 U; 0 Other;

Query Match        100.0%;  Score 5676;  DB 5;  Length 5960;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 5676;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 CAGCTGTTCAGGGATGTCTCGAAAAGAGCCACCCACATTGCTTCTGGACACTGGGTGT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CAGCTGTTCAGGGATGTCTCGAAAAGAGCCACCCACATTGCTTCTGGACACTGGGTGT 60

Qy         61 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACTTTGGAGGGTATCAGGTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAAT 120

Qy        121 TGGCCCTCGTTCCCTTGCATGCCCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGGCCCTCGTTCCCTTGCATGCCCTTCCTTGGGACACTCCCTTAAGGCATCTTCTTG 180

Qy        181 ACATTAACTAACTATAAAATGTTTATTGATGAATTTCAGTGACCTGAAGAGAGTGGAG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACATTAACTAACTATAAAATGTTTATTGATGAATTTCAGTGACCTGAAGAGAGTGGAG 240

Qy        241 GTCAAATCAGAAGAAGCACATGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTCAAATCAGAAGAAGCACATGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAA 300

Qy        301 AGTCATTCGAATACCATTCAGTTACTTACTAAGTTCTAGGCCCAGCTTACTCCTAATCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGTCATTCGAATACCATTCAGTTACTTACTAAGTTCTAGGCCCAGCTTACTCCTAATCGAT 360

Qy        361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGACAGGATCCTCTCCATGA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCAGACTGTAGCAAATATTAGGTCCAAAGTTGGAAGAGTTAGACAGGATCCTCTCCATGA 420

Qy        421 CAGAACTTTGGCTTCCACTTTACTAGAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CAGAACTTTGGCTTCCACTTTACTAGAAATAGAGATTGTGTGGTTGAGCTGCAGCTATGTA 480

Qy        481 CAGAAAAGTGTCATACAAGTTTAAAAATCACCAAACTCAGTCTTCTTCAATTTGAGCAATAGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGAAAAGTGTCATACAAGTTTAAAAATCACCAAACTCAGTCTTCTTCAATTTGAGCAATAGT 540

Qy        541 TGGTGAATTTACTCCACCACCTCGAAGGTTCTTTCCTGCTCTCCTCACTATA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGGTGAATTTACTCCACCACCTCGAAGGTTCTTTCCTGCTCTCCTCACTATA 600

Qy        601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTGTTCAGTTCAGTTCAGACAACAAGGAAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AATGCAGGATGACCTGGAAAGGCTAGGACCTGAGGTTCAGTTGTTCAGTTCAGTTCAGACAACAAGGAAT 660

Qy        661 TCAGTTTTCTCGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TCAGTTTTCTCGATCTCATAGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCAT 720

Qy        721 CTGCTTAAGCCTCTTGTGGTGGCATCTGTTGTTTCCACTGCCCGTGACCTATTGCTCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTGCTTAAGCCTCTTGTGGTGGCATCTGTTGTTTCCACTGCCCGTGACCTATTGCTCT 780

Qy        781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTTCGCTCTGCCAGTCATGGTAGGG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTCCTTGGTTAACAGAACCTTTATTTTCTTCTGAAAACTCTTCGCTCTGCCAGTCATGGTAGGG 840

Qy        841 CCATCAGTCCACATGATCAGGCCCTTCCTGCCCAGCCAAACATGGCATCTTTCTTTGGGAATT 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCATCAGTCCACATGATCAGGCCCTTCCTGCCCAGCCAAACATGGCATCTTTCTTTGGGAATT 900

Qy        901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTAC 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TGAATCTTAAGCTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTAC 960

Qy        961 AGTGGCTTGCAAAGTGACTGCTCCATTCCTTAAGTCCCTGAATTTATAAATTTAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGTGGCTTGCAAAGTGACTGCTCCATTCCTTAAGTCCCTGAATTTATAAATTTAT 1020

Qy       1021 CCTGGTTACAGCCCTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTATAGTCTGT 1080
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCTGGTTACAGCCCTTCTGAGATGTGTGGTTTTTTTCCAACTGTCTCTATAGTCTGT 1080

Qy       1081 GAATTTCATATTTCTTTCATACATTTCATGTTTTGTTTGTTTGTGTGTTTTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAATTTCATATTTCTTTCATACATTTCATGTTTTGTTTGTTTGTGTGTTTTTT 1140

Qy       1141 GGCTTTAGGTAGGCAGAATCAGTTTCGTGTTTATACCCAAGGAATCCTGATTGATACA 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCTTTAGGTAGGCAGAATCAGTTTCGTGTTTATACCCAAGGAATCCTGATTGATACA 1200

Qy       1201 TCCTTCCCTTAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCCTTCCCTTAAAATAAAGTATCTAAGGCTCAAAGAGAGTAGGCTACCTGCCTGAGG 1260

Qy       1261 TCTGGGAGTAAGTTAGTACCCAGAGCTCGTACTAACCCCAGGTTAGCCAACTGCTTTACAC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TCTGGGAGTAAGTTAGTACCCAGAGCTCGTACTAACCCCAGGTTAGCCAACTGCTTTACAC 1320

Qy       1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAGAGAGCTACTATTTTGC 1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AACATTTGCTCTCTCCTTCAGAGTTATAGCAGTCTTGGAAGAAGAGAGCTACTATTTTGC 1380

Qy       1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA 1440
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CAAAGACCTCAGGAGGACCAAGAACAAGTTCTGGGATATGTGATGATTGAACTCTTAAAA 1440

Qy       1441 AGTTGTTGGACTTCTGGCCATAATTGTGATCTAAGACCAGATTTCATTCTTCTAAATAGCT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AGTTGTTGGACTTCTGGCCATAATTGTGATCTAAGACCAGATTTCATTCTTCTAAATAGCT 1500

Qy       1501 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC 1560
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAACAAACAAACAAGAGATCCACAGGTTCAGCAGTTCAGCAGCTATAATAAGAGTGAATTACTGATAC 1560

Qy       1561 AGTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 AGTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAAATCCAGAC 1620

Qy       1621 ACAGAGAATACGTACCACCATATGCCTGCATTATGTGATATTCGAGCATTGTGATTGTCCAA 1680
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ACAGAGAATACGTACCACCATATGCCTGCATTATGTGATATTCGAGCATTGTGATTGTCCAA 1680

Qy       1681 CATATTAGCCAGTGCTCTTCAGTAGCGTTAGCCACACATCAGAC 1740
              |||||||||||||||||||||||||||||||||||||||||||
Db       1681 CATATTAGCCAGTGCTCTTCAGTAGCGTTAGCCACACATCAGAC 1740

Qy       1741 ATAAAAATTTAGCTCTTATTATAGAATATTTCCTTTATAACGAAAGTTCTATTGGAAAACAATGTTCTAG 1800
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 ATAAAAATTTAGCTCTTATTATAGAATATTTCCTTTATAACGAAAGTTCTATTGGAAAACAATGTTCTAG 1800

Qy       1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG 1860
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GGTGCAAATATAGAATATTTCCTTTATAACAGAAAGTTCTATTGGAAAACAATGTTCTAG 1860

Qy       1861 AAAATATACACATAATCATATAAAACAAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG 1920
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 AAAATATACACATAATCATATAAAACAAAAAGCAAGTCAGTGATTGTCTAAGGCCAGGGG 1920
```

us-10-030-294-1.rng

Db   4081 GCAATTAACAAATCTATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTTG 4140

Qy   4141 TGGAGCAAGCTGTTTGATTTGGCTGGGGGCTCAGCCTGTTGTGAATTTCACAATT 4200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 TGGAGCAAGCTGTTTGATTTGGCTGGGGGCTCAGCCTGTTGTGAATTTCACAATT 4200

Qy   4201 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAGAATGTCAAGTTTAAATAGCT 4260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 CACAGATGTTAGCCGCTCTCGGGCTAAGTAAAGGAAGAGAATGTCAAGTTTAAATAGCT 4260

Qy   4261 TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTATGAAACACATTTTGAGT 4320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 TCTCCCTTCCATCCTGGCTGAAGCAACAAATAAAATATTTTATGAAACACATTTTGAGT 4320

Qy   4321 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 4380
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 TAGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAAC 4380

Qy   4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGGTGAAGGGA 4440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TTTGACATCTACTGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGGGTGAAGGGA 4440

Qy   4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTCCCCTG 4500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AGATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTCCCCTG 4500

Qy   4501 TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCACAAA 4560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TTGGAGTCTATCCTAACTGAGCTTCTGAATCATATTTCATTCAAATTCCACAAA 4560

Qy   4561 ACCAGGATAAGTTTACAGCCCATATTCAGAAAGGAAATAAATTATTTTGTGTGTAGACTT 4620
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ACCAGGATAAGTTTACAGCCCATATTCAGAAAGGAAATAAATTATTTTGTGTGTAGACTT 4620

Qy   4621 TCCTGATATTACACTGATTTGGAATATATGAACAATTTTATGGTTTCCTTCGAAGTAG 4680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 TCCTGATATTACACTGATTTGGAATATATGAACAATTTTATGGTTTCCTTCGAAGTAG 4680

Qy   4681 GTCAAGTCAAAGCAAACCAAAACAGCAAAAACTGTAAGCACATAAAGAATAGAGTGGAG 4740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 GTCAAGTCAAAGCAAACCAAAACAGCAAAAACTGTAAGCACATAAAGAATAGAGTGGAG 4740

Qy   4741 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 4800
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 CCGACTGAGAGATTAAAATAAACTAGAATATTTTATTAACAGGCAATTTGAAATAATTT 4800

Qy   4801 GTGCACTTCAGAAATATTCTACAATAATAATATTTCCAATTTAATATCTTTAAGAAAA 4860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 GTGCACTTCAGAAATATTCTACAATAATAATATTTCCAATTTAATATCTTTAAGAAAA 4860

Qy   4861 TTACTATATTATATGTAAGTAACATGTGCATGTGTTTGAGGTAGTAGAATTAACTCAATAA 4920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 TTACTATATTATATGTAAGTAACATGTGCATGTGTTTGAGGTAGTAGAATTAACTCAATAA 4920

Qy   4921 AGGTTATTTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGGATATCT 4980
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 AGGTTATTTTCTTTATTCGGGTCAGGCAAAGCTTCTAAGGGGATGTGAAAGGGGATATCT 4980

Qy   4981 CTTTCTCTTAGCTGGAGGAAGAGTTCTAAGTTCTAAGTTAAATAATAATCAAGGAATTTCCCT 5040
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 CTTTCTCTTAGCTGGAGGAAGAGTTCTAAGTTCTAAGTTAAATAATAATCAAGGAATTTCCCT 5040

Qy   5041 GTCTTTGCTATTTGACCTGAGATTGTGACCACCAACAGCCGTTGGCTGAAAGGGAAACTGAAGGG 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 GTCTTTGCTATTTGACCTGAGATTGTGACCACCAACAGCCGTTGGCTGAAAGGGAAACTGAAGGG 5100

Qy   5101 CGGGGAGGGAGGGAAATAGATAGAATGAAAAAAACAAAACAAAAACAAAACTTCCCTAAGCAGCTCT 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 CGGGGAGGGAGGGAAATAGATAGAATGAAAAAAACAAAACAAAAACAAAACTTCCCTAAGCAGCTCT 5160

Qy   5161 ACAAAACATTTTAGCCCAGAGAAATAGTCACAGAAATCCTCAAATCCAGTATCCAGA 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 ACAAAACATTTTAGCCCAGAGAAATAGTCACAGAAATCCTCAAATCCAGTATCCAGA 5220

Qy   5221 TACAAGGAAGTGTTATGTAGCTGGAGCAGCTGGACACTCATCAGTTCAGTTACA 5280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 TACAAGGAAGTGTTATGTAGCTGGAGCAGCTGGACACTCATCAGTTCAGTTACA 5280

Qy   5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCATCCAATCACGACA 5340
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 AAAGTCCAGGCTGCTGAAATTAAACTCTGATGCCATTCATGCCAGCATCCAATCACGACA 5340

Qy   5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGGCTACTA 5400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 GAGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGGCTACTA 5400

Qy   5401 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGGAGGATGCCCA 5460
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TACGTCAAGGACTCTGAAGCCGTGAGAGAGGGGAAGAACAACAGTAGAGGAGGATGCCCA 5460

Qy   5461 GCTGGTAAGAATCGAGTGTTTAGTGAATTTTAGTCAATTCATTGAATCTCATTGGCTAAAA 5520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 GCTGGTAAGAATCGAGTGTTTAGTGAATTTTAGTCAATTCATTGAATCTCATTGGCTAAAA 5520

Qy   5521 TCAAGAAACGCTCCGCCTCTTGCAAATATGTATGAAGGAGGAGAAGTGCCTAAACTTCTA 5580
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 TCAAGAAACGCTCCGCCTCTTGCAAATATGTATGAAGGAGGAGAAGTGCCTAAACTTCTA 5580

Qy   5581 TGTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCGGCTTTATATCTATATATACAC 5640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TGTCTGATAGCATTTGACCCTATTGCTTTTAGCCTCCGGCTTTATATCTATATATACAC 5640

Qy   5641 AGGTATTTGTGTATATTTTTATATATTTTTATATATAATTGTTCTCCGT 5676
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 AGGTATTTGTGTATATTTTTATATATTTTTATATATAATTGTTCTCCGT 5676

RESULT 5
AAF25915
ID   AAF25915 standard; DNA; 13940 BP.
XX
AC   AAF25915;
XX
DT   19-APR-2001 (first entry)
XX
DE   Human tumor suppressor gene p51 promoter DNA fragment SEQ ID 12.
XX
KW   Tumor suppressor; p51; cell death; cell proliferation; cancer;
KW   cytostatic; gene therapy; screening; ds.
XX
OS   Homo sapiens.
XX
PN   WO200100818-A1.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-JP004261.
XX
PR   29-JUN-1999; 99JP-00183195.
XX
PA   (NIPK) NIPPON KAYAKU KK.
PA   (SAKA/) SAKAI T.
XX
PI   Sakai T, Kagaya S, Sato T, Sukenaga Y, Fujii H;
XX
DR   WPI; 2001-112452/12.
XX
PT   New DNA sequences encoding the human p51 promoter domain for screening
PT   for compounds that modify p51 promoter activity and for diagnosing and
PT   treating cancer.
XX
PS   Example 1; Page 49-57; 60pp; Japanese.
XX
CC   This invention describes novel DNA sequences (I) encoding the human p51
CC   promoter domain (which may include the 5'-untranslated sequence) or
CC   derived from it by addition, deletion and/or substitution of one or more

```
CC  bases. The invention also describes (1) expression plasmids including (I)
CC  ; (2) host cells transformed by (1); (3) DNA probes binding to all or
CC  part of (I); (4) cloning (I) using (3); (5) DNA (and RNA corresponding to
CC  it) inhibiting the expression of p51, which includes the antisense
CC  sequence of (I); (6) screening compounds for their ability to modify p51
CC  promoter activity, by observing their effect on cells transformed by (I);
CC  (7) compounds identified by (6); and (8) drug compositions containing
CC  (7). The products of the invention have cytostatic activity and can be
CC  used for gene therapy. (I) is used to screen for compounds that modify
CC  p51 promoter activity. (I) and RNA corresponding to it are used to
CC  inhibit the expression of p51. (I) and the identified compounds are used
CC  for the diagnosis and treatment of cancer
XX
SQ  Sequence 13940 BP; 3949 A; 2972 C; 3085 G; 3934 T; 0 U; 0 Other;

Query Match             100.0%;  Score 5673.4;  DB 5;  Length 13940;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 5674;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   0;
```

```
Qy      2  AGCTGTTCAGGGATGCTCTGGAAAGAAGCCACCCACATTGCTTCTGGACACTGGGTGTG   61
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51  ATCTGTTCAGGGATGCTCTGGAAAGAAGCCACCCACATTGCTTCTGGACACTGGGTGTG  110

Qy     62  ACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAATT  121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111  ACTTTGGAGGGTATCAGGTTTGTCTGTTAAAGAAACTGCCAACCTCTTCCTGCCCCAATT  170

Qy    122  GGCCCTCGTTCCCTGCATGCCCTCTTCCTTGGGACACTCTTCTTGA  181
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    171  GGCCCTCGTTCCCTGCATGCCCTCTTCCTTGGGACACTCTTCTTGA  230

Qy    182  CATTAACTTAACTATAAATGTTTATTTGATGAATTTGATGAATTTCAGTGACCTGAAGATGGAGG  241
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231  CATTAACTTAACTATAAATGTTTATTTGATGAATTTGATGAATTTCAGTGACCTGAAGATGGAGG  290

Qy    242  TCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAAA  301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291  TCAAATCAGAAGAAGCACATGGCTAAGGTTGCAATGCACTTGCTTTTTCATTGAATTAAA  350

Qy    302  GTCATTGAATACCATTCAGTTACTTAAGTTCTAGGCCCAGCTTACTTGAGCAATAGTT  361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351  GTCATTGAATACCATTCAGTTACTTAAGTTCTAGGCCCAGCTTACTTGAGCAATAGTT  410

Qy    362  TCAGACTGTAGCAAATATTAGGTCGGAAGAGTTAGCAGGATCCTCTCCATGAC  421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411  TCAGACTGTAGCAAATATTAGGTCGGAAGAGTTAGCAGGATCCTCTCCATGAC  470

Qy    422  AGAACTTGGCTTCCACTTACTAAGTTGTGGTTGAGCTGAGCTATGTAC  481
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    471  AGAACTTGGCTTCCACTTACTAAGTTGTGGTTGAGCTGAGCTATGTAC  530

Qy    482  AGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTTTGAGCAATAGTT  541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531  AGAAAAGTGTCATACAATTAAAAATCACCAAACTCAGTCTTTGAGCAATAGTT  590

Qy    542  GGTGAATTTACTCCACCACCTCTCTCCTGAAGGTTCTTTCCTGCTCCTCTCCTACTATAA  601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591  GGTGAATTTACTCCACCACCTCTCTCCTGAAGGTTCTTTCCTGCTCCTCTCCTACTATAA  650

Qy    602  ATGCAGGATGACCTTGGAAAGGCTAGGAACCTGAGTTCAGTTACCCTGACCAAAAGGAATT  661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651  ATGCAGGATGACCTTGGAAAGGCTAGGAACCTGAGTTCAGTTACCCTGACCAAAAGGAATT  710

Qy    662  CAGTTTCTCTGATCTCATAGTCACAGGCTCTGCCAGAGCTCACGGAACATGCAAGATCATC  721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711  CAGTTTCTCTGATCTCATAGTCACAGGCTCTGCCAGAGCTCACGGAACATGCAAGATCATC  770

Qy    722  TGCTTTAAGCCTCTGTGGTGGCATCGTGTTTTTCCACTGCCCTGTACCTATTGCTCTT  781
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771  TGCTTTAAGCCTCTGTGGTGGCATCGTGTTTTTCCACTGCCCTGTACCTATTGCTCTT  830

Qy    782  TCCTTGGTTAACAGACCTTTATTTCTTCTGAAAACTCTCTGCTCAGTCATGGTAGGGC  841
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831  TCCTTGGTTAACAGACCTTTATTTCTTCTGAAAACTCTCTGCTCAGTCATGGTAGGGC  890

Qy    842  CATCAGTCCACATGATCAGGCCTCCTGCCCAAACATGGCATCTTCTTTTGGGAATTT  901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891  CATCAGTCCACATGATCAGGCCTCCTGCCCAAACATGGCATCTTTCTTTTGGGAATTT  950

Qy    902  GAATCTTAAGCTGAATAGCTGAAGTGAAGTTCAAAAAAAGCTGTGAATCTGACTTACGCCTACA  961
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951  GAATCTTAAGCTGAATAGCTGAAGTGAAGTTCAAAAAAAGCTGTGAATCTGACTTACGCCTACA  1010

Qy    962  GTGGCTTGCAAAGTGACTGTCCATTCCTATTCCTATTCCTTAAGTCCCTGAATTTATAATTTATC  1021
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011  GTGGCTTGCAAAGTGACTGTCCATTCCTATTCCTATTCCTTAAGTCCCTGAATTTATAATTTATC  1070

Qy   1022  CTGGTTACAGCCCTTTCTGAGATGTGGTTTTTTCCAACTGTCTCTTATAGTCTGTG  1081
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1071  CTGGTTACAGCCCTTTCTGAGATGTGGGTTTTTTTCCAACTGTCTCTTATAGTCTGTG  1130

Qy   1082  AATTTCATATTCTTTCATACATTTCATGTTTTGTTTGTTTGTTTGTGTGTGTTTTTG  1141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131  AATTTCATATTCTTTCATACATTTCATGTTTTGTTTGTTTGTTTGTGTGTGTTTTTG  1190

Qy   1142  GCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACAT  1201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191  GCTTTAGGTAGGCAGAATCAGTTTCTGTTGTTTATACCCAAGGAATCCTGATTGATACAT  1250

Qy   1202  CCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGGT  1261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251  CCTTCCCCTTAAAAATAAAGTATCTAAGGCTCAAAGAGTAGGCTACCTGCCTGAGGT  1310

Qy   1262  CTGGGAGTAAGTTAGTACCAGAGCTCGTACTCAACCCCAGGTTAGCCAACTGCTTTACACA  1321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311  CTGGGAGTAAGTTAGTACCAGAGCTCGTACTCAACCCCAGGTTAGCCAACTGCTTTACACA  1370

Qy   1322  ACATTTGCTCCTCTCCTTCAGAGTTATAGACAGTCTTGGAGAAAAGAAGCTACTATTTGCC  1381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371  ACATTTGCTCCTCTCCTTCAGAGTTATAGACAGTCTTGGAGAAAAGAAGCTACTATTTGCC  1430

Qy   1382  AAAGACCTCAGGAGGACCAAGAACAAGTTCTGGAATGTGATGATTGAACTCTTAAAAA  1441
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431  AAAGACCTCAGGAGGACCAAGAACAAGTTCTGGAATGTGATGATTGAACTCTTAAAAA  1490

Qy   1442  GTTTGTTGGACTTCTGGCCATAATTGTGTATCTAAGACCAGATTCATTCTTAATAGCTA  1501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491  GTTTGTTGGACTTCTGGCCATAATTGTGTATCTAAGACCAGATTCATTCTTAATAGCTA  1550

Qy   1502  AACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATACA  1561
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551  AACAAACAAACAAGAGATCCACAGGTTCAGCAGCTATAATAAGAGTGAATTACTGATACA  1610

Qy   1562  GTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAATCCAGACA  1621
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611  GTTGACAACATGAATATCTCAGAAACCATGGCATCAATGAGCAAAAAATCCAGACA  1670

Qy   1622  CAGAAGAATACGTACCCATATGCCTGCATTTATGTGATATTCTAGACATTGTTCCAAC  1681
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1671  CAGAAGAATACGTACCCATATGCCTGCATTTATGTGATATTCTAGACATTGTTCCAAC  1730

Qy   1682  ATAGTAGCCAGTAGCCTCACATGGCTATCCAAATTTAAGTTGATTAAAATTAAGTAAGAA  1741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1731  ATAGTAGCCAGTAGCCTCACATGGCTATCCAAATTTAAGTTGATTAAAATTAAGTAAGAA  1790

Qy   1742  TAAAAATTTAGCTCTTCAGTAGGCTTACGGAACATGCAAGATCATC  1801
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   1791  TAAAAATTTAGCTCTTCAGTAGGCTTACGGAACATGCAAGATCATC  1850

Qy   1802  GTGCAAATATAGAAATATTCCTTTATAAACAAAAAACAATGTTCTAGA  1861
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1851  GTGCAAATATAGAAATATTTCCTTTATAAACAAAAAACAATGTTCTAGA  1910

Qy   1862  AAATATACACATAATCTATAAAAACAAAAACAAAAAAACAAGTCAGTGATTGCTCAAGGCCAGGGGT  1921
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911  AAATATACACATAATCTATAAAAACAAAAACAAAAAAACAAGTCAGTGATTGCTCAAGGCCAGGGGT  1970
```

Sequence alignment (Qy = query, Db = database). Vertical bars indicate base identity between the two sequences.

```
Qy 1922 GAGGGGAGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGGTAATAGGGTTGTTGGAA 1981
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1971 GAGGGGAGATCGATTGCAAAGTGGTATGAGGAAAGTTTGGGGGTAATAGGGTTGTTGGAA 2030

Qy 1982 TCTTGATTGCGATGAGGCTACTCGGTGTCTAAGTGTCACCTCAGACTGAACACTT 2041
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2031 TCTTGATTGCGATGAGGCTACTCGGTGTCTAAGTGTCACCTCAGACTGAACACTT 2090

Qy 2042 GGAATTGGCGAATTCATTGTGACCAATTTATACCTCATAAAGTAACTCTAAGAGGTCAA 2101
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2091 GGAATTGGCGAATTCATTGTGACCAATTTATACCTCATAAAGTAACTCTAAGAGGTCAA 2150

Qy 2102 GTGTTTGTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTGCAA 2161
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2151 GTGTTTGTGTGGAAATTATTTTAATCAGTTGCAATACTTATTATGAGATGATTTTGCAA 2210

Qy 2162 ATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTCCTGAAAACACA 2221
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2211 ATACATAAACATGTTATTCATCCATTAGGTGCAATATTTTGCTAGCTCCTGAAAACACA 2270

Qy 2222 GAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATGA 2281
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2271 GAGATGAATTAGAATAGCAAGCCTGCCCTCAAGCTGTTCACAATCCAGTACAGGAGATGA 2330

Qy 2282 GTCTATTCAAAAAATAGCTAGACTCCAGGAGGAAAGTTATAGGTGACCTTACACAAAAAG 2341
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2331 GTCTATTCAAAAAATAGCTAGACTCCAGGAGGAAAGTTATAGGTGACCTTACACAAAAAG 2390

Qy 2342 TGCAGATATAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTTATGTGAAAAAAG 2401
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2391 TGCAGATATAATTATGTAGGACAGTAGAAGTGGGGAAGGTTTCTTTTATGTGAAAAAAG 2450

Qy 2402 AGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGACAGATGGAGTTTT 2461
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2451 AGGGAGAATTTTTGGTCTTTGAAGGATGAGCAAGATGTGAATATGCGACAGATGGAGTTTT 2510

Qy 2462 AAAACATTCCTGGTCGTGGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATAT 2521
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2511 AAAACATTCCTGGTCGTGGAGGGCAGAATATGATCCAAGGCACAAGAGCAACCAGAAAAATAT 2570

Qy 2522 GCAACCTAGAGGGAAAGTCATGATGAAAGTGCAGTTGTAAAATAATTTTCATGAATGTAAG 2581
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2571 GCAACCTAGAGGGAAAGTCATGATGAAAGTGCAGTTGTAAAATAATTTTCATGAATGTAAG 2630

Qy 2582 TGAGAAGAATTTGTATCATAGACACCTGAGTTGGCAGAGTGCATGTTCTTGGCGTTCCTAG 2641
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2631 TGAGAAGAATTTGTATCATAGACACCTGAGTTGGCAGAGTGCATGTTCTTGGCGTTCCTAG 2690

Qy 2642 GAGTCAAGAGAACAAAGTGTCCCCTTTCTCCTACGTTATGCTCAGTGGTCCAAGTCCAAA 2701
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2691 GAGTCAAGAGAACAAAGTGTCCCCTTTCTCCTACGTTATGCTCAGTGGTCCAAGTCCAAA 2750

Qy 2702 ACACCTTTCCTTCCTTAAGTACTTCTTCTTCTCCTCCCCCATACAAATCTAAAGTCTTCACAA 2761
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2751 ACACCTTTCCTTCCTTAAGTACTTCTTCTTCTCCTCCCCCATACAAATCTAAAGTCTTCACAA 2810

Qy 2762 ACATCATTTAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTTTCCTAGAACTTCTATG 2821
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2811 ACATCATTTAAACAGGCAGGTCATGGTCAGAAAGGCAATTGCTTTTTCCTAGAACTTCTATG 2870

Qy 2822 TACGTTATTATTACAATTTCTGCGCTAGTAGTAACCTTAAAGTTCCACC 2881
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 2871 TACGTTATTATTACAATTTCTGCGCTAGTAGTAACCTTAAAGTTCCACC 2930

Qy 2882 TTGCACATCAAAGATATAATTCATGCATTTGTATAGTAGTAACCTTAGTCCCCTAAGAGAATA 2941
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2931 TTGCACATCAAAGATATAATTCATGCATTTGTATAGTAGTAACCTTAGTCCCCTAAGAGAATA 2990

Qy 2942 AGGATGAACTATAAAATATAAGAAGTAATTATGGTAATTATATATATGATTGCCACTTATTT 3001
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2991 AGGATGAACTATAAAATATAAGAAGTAATTATGGTAATTATATATATGATTGCCACTTATTT 3050

Qy 3002 TTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTCTGTTGTCTGTGAATTCTAGAGAGTTTGC 3061
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3051 TTCACTTGATCGTGTATGGTTGCATGCTACTGGTGTCTGTTGTCTGTGAATTCTAGAGAGTTTGC 3110

Qy 3062 CTCTTTTCCTGGGTCAACTCTCGCCATTATTCCATAATGCAATGCCAATCTTT 3121
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3111 CTCTTTTCCTGGGTCAACTCTCGCCATTATTCCATAATGCAATGCCAATCTTT 3170

Qy 3122 TTCATAATTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCCTTAGTAA 3181
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3171 TTCATAATTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCCTTAGTAA 3230

Qy 3182 CTTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATATTT 3241
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3231 CTTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATATTT 3290

Qy 3242 AAATACGTTAGGCTATAAACCTTTTAAAATTTTTAAAAAAATAGATGAGTGGTGGTGGCTC 3301
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3291 AAATACGTTAGGCTATAAACCTTTTAAAATTTTTAAAAAAATAGATGAGTGGTGGTGGCTC 3350

Qy 3302 ATGCCTGTAATCCCACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGT 3361
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3351 ATGCCTGTAATCCCACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGT 3410

Qy 3362 TTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAACAAAACAAAACAAA 3421
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3411 TTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAACAAAACAAAACAAA 3470

Qy 3422 ATTACCTGGGTATGGTTGTGTCAGTCCAAGCTACCAGGAGCTACACAGGAGGCAGAA 3481
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3471 ATTACCTGGGTATGGTTGTGTCAGTCCAAGCTACCAGGAGCTACACAGGAGGCAGAA 3530

Qy 3482 GGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGGCGTCTACACTCA 3541
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3531 GGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGGCGTCTACACTCA 3590

Qy 3542 GTCTGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAATAAAAATAAACTTTTA 3601
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3591 GTCTGGGTGACAGTGCAAGAGCTGTCTCAAAAATAATAAATAAAAATAAACTTTTA 3650

Qy 3602 AAAACAAAAATTAATTAAATTTTAAAAACAACAACACTAGAGATGTTTGCAAATTGAT 3661
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3651 AAAACAAAAATTAATTAAATTTTAAAAACAACAACACTAGAGATGTTTGCAAATTGAT 3710

Qy 3662 TATTTGGGAGTCTATATCCCTGGAAGTTAATTTAAAATATTTAGAAGAGTTCTTCCTCAT 3721
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3711 TATTTGGGAGTCTATATCCCTGGAAGTTAATTTAAAATATTTAGAAGAGTTCTTCCTCAT 3770

Qy 3722 TTCCTAGAGACGTCGAATTGTAAATATCAGAGCTGAAGAACACTAGGGCTCGCCACTC 3781
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3771 TTCCTAGAGACGTCGAATTGTAAATATCAGAGCTGAAGAACACTAGGGCTCGCCACTC 3830

Qy 3782 CAAAGTGTGGTCCAAGGACCAGCAGCATCGGAACCTGGAACGTGTAGAAATGCAGA 3841
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3831 CAAAGTGTGGTCCAAGGACCAGCAGCATCGGAACCTGGAACGTGTAGAAATGCAGA 3890

Qy 3842 GTCTTAGGCCTTCACCCCCAGACCTACTGCATTAACAGAATCGCATTAAACAAGATTTCTAGGTG 3901
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3891 GTCTTAGGCCTTCACCCCCAGACCTACTGCATTAACAGAATCGCATTAAACAAGATTTCTAGGTG 3950

Qy 3902 CCTTCACCCGGCACATTAAAACTTGGGATGGTGTGCTGTCACGGAATTTTAATTTTCAGAAC 3961
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3951 CCTTCACCCGGCACATTAAAACTTGGGATGGTGTGCTGTCACGGAATTTTAATTTTCAGAAC 4010

Qy 3962 TTATAAATGGAATCACTTGGGCTGTGGCTGTGTCACAGGAAATTGATTATTTTTAATTTCAGAAC 4021
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4011 TTATAAATGGAATCACTTGGGCTGTGGCTGTGTCACAGGAAATTGATTATTTTTAATTTCAGAAC 4070

Qy 4022 CTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGGAAGAAGCCAAACTCTTTAACTG 4081
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4071 CTTCTATTTAGGTCATCTATATTTGCTAATAGCAGGGAAGAAGCCAAACTCTTTAACTG 4130

Qy 4082 CAATTAACAAATCTATAATTAATTAAGTTAAGCAATCTTCCCTTAAGTTTTACATTTTGT 4141
```

Db 4131 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CAATTAACAAATCTATATAATTAGTTAAGCAATCTTCCCTTTACATTTTGT 4190

Qy 4142 GGAGCAAGCTGTTTGATTTGGCTCGGGCTCAGGCCGCCTGTTTGTGAATTTCACAATTC 4201
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4191 GGAGCAAGCTGTTTGATTTGGCTCGGGCTCAGGCCGCCTGTTTGTGAATTTCACAATTC 4250

Qy 4202 ACAGATGTTAGCCGCTCTCCGGCTAAGTAAGGAGGAAGAGAATGTCAAGTTTTAAATAGCTT 4261
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4251 ACAGATGTTAGCCGCTCTCCGGCTAAGTAAGGAGGAAGAGAATGTCAAGTTTTAAATAGCTT 4310

Qy 4262 CTCCCCTTCCATCCTGCTGAAGCAACAAATAAAATATTTTATGAAAACACATTTTGAGTT 4321
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4311 CTCCCCTTCCATCCTGCTGAAGCAACAAATAAAATATTTTATGAAAACACATTTTGAGTT 4370

Qy 4322 AGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAACT 4381
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4371 AGATTTACTTACAGGGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCACAACT 4430

Qy 4382 TTGACATCTCGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGGAA 4441
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4431 TTGACATCTCGATGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGGAA 4490

Qy 4442 GATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTCCCCTGT 4501
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4491 GATGTGAACTCACCATGTTAGTGACCGTTAGATACACAGAGTGGTTTTTTTCCCCTGT 4550

Qy 4502 TGGAGTCTATCCTAACTGAGCTTCTGAAATCATATTTCATTCAATTTCCAAATCCACAAAA 4561
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4551 TGGAGTCTATCCTAACTGAGCTTCTGAAATCATATTTCATTCAATTTCCAAATCCACAAAA 4610

Qy 4562 CCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATATTTGTGTGTAGACTTT 4621
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4611 CCAGGATAAGTTACAGCCCATATTCAGAAAGGAAATAAATATTTGTGTGTAGACTTT 4670

Qy 4622 CCTGATATTACACTGATTTGGAATATATGAACAATTTATGGTTTCCTTTCGAAGTAGG 4681
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4671 CCTGATATTACACTGATTTGGAATATATGAACAATTTATGGTTTCCTTTCGAAGTAGG 4730

Qy 4682 TCAAGTCAAAGCAAAACCAAAAACAGCAAAAACTGTAAGACATAAAGAATAGAGTGGAGC 4741
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4731 TCAAGTCAAAGCAAAACCAAAAACAGCAAAAACTGTAAGACATAAAGAATAGAGTGGAGC 4790

Qy 4742 CGACTGAGAGATTAAAATAAACTAGAATATTTTTATTAACAGGCAATTTGAAATAATTTG 4801
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4791 CGACTGAGAGATTAAAATAAACTAGAATATTTTTATTAACAGGCAATTTGAAATAATTTG 4850

Qy 4802 TGCACTTCAGAATATTCTACAATATATATTATTATTCCAATTTAATATCTTAAGAAAAT 4861
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4851 TGCACTTCAGAATATTCTACAATATATATTATTATTCCAATTTAATATCTTAAGAAAAT 4910

Qy 4862 TACTATATATATGTAAGTACATGTCATGTGTTGAGGTAGGAGATATTTAACTCAATAAA 4921
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4911 TACTATATATATGTAAGTACATGTCATGTGTTGAGGTAGGAGATATTTAACTCAATAAA 4970

Qy 4922 GGTTATTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGATGTGAAAGGGATATCTC 4981
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4971 GGTTATTTCTTTTATTCGGGTCAGGCAAAGCTTCTAAGGGATGTGAAAGGGATATCTC 5030

Qy 4982 TTTCTCCTTAGCTGCGAGGAGAGTGAGTTCTAAGTAAGTTAAATAATCAAGGAATTCCCTG 5041
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5031 TTTCTCCTTAGCTGCGAGGAGAGTGAGTTCTAAGTAAGTTAAATAATCAAGGAATTCCCTG 5090

Qy 5042 TCTTTGCTATTTGAGATTGTGACCACCAACAGGCCGTTGGCTGAAAGGGAAACTGAAGGGC 5101
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5091 TCTTTGCTATTTGAGATTGTGACCACCAACAGGCCGTTGGCTGAAAGGGAAACTGAAGGGC 5150

Qy 5102 GGGGAGGGGAGGGAAATAGATGAAAACAAAAACAAAACTTCCCTAAGCAGCTCTA 5161
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5151 GGGGAGGGGAGGGAAATAGATGAAAACAAAAACAAAACTTCCCTAAGCAGCTCTA 5210

Qy 5162 CAAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGAT 5221
       |||||||||||||||||||||||||||||||||||||||||||||||||||

Db 5211 CAAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGAT 5270

Qy 5222 ACAAGGAAGTGTTATGTAGCTGGAGCAGGTGGACTCATCAGGTCAGTTCAGTTACAA 5281
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5271 ACAAGGAAGTGTTATGTAGCTGGAGCAGGTGGACTCATCAGGTCAGTTCAGTTACAA 5330

Qy 5282 AAGTCCAGGCTGCTGAAATTAAACTCGATGCCATTCATGCCAGCATCCAATCACGACAG 5341
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5331 AAGTCCAGGCTGCTGAAATTAAACTCGATGCCATTCATGCCAGCATCCAATCACGACAG 5390

Qy 5342 AGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACACAGTGGCTACTAT 5401
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5391 AGATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACACAGTGGCTACTAT 5450

Qy 5402 ACGTCAAGGACTCTGAAGCCGTGAGAGCCGTGAGAGAGGGGGAAGAACACAGTAGAGAGGATGCCCAG 5461
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5451 ACGTCAAGGACTCTGAAGCCGTGAGAGCCGTGAGAGAGGGGGAAGAACACAGTAGAGAGGATGCCCAG 5510

Qy 5462 CTGGTAAGAATCAGAGTGTTTAGTAGAGTTTAGTCAATTGATGAATTCATTGGCTAAAAT 5521
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5511 CTGGTAAGAATCAGAGTGTTTAGTAGAGTTTAGTCAATTGATGAATTCATTGGCTAAAAT 5570

Qy 5522 CAAGAAACGCTCCGGCCTCTTTGCAAATATGTATGAAAGGAGGAGGAGTGCCTAAAACTTCTAT 5581
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5571 CAAGAAACGCTCCGGCCTCTTTGCAAATATGTATGAAAGGAGGAGGAGTGCCTAAAACTTCTAT 5630

Qy 5582 GTCTGATAGGCATTTGACCCCTATTGCTTTAGCCTTCTAATCTATATATACACA 5641
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5631 GTCTGATAGGCATTTGACCCCTATTGCTTTAGCCTTCTAATCTATATATACACA 5690

Qy 5642 GGTATTTGTGTATATTTATATAAATTGTTCTCCGT 5676
       |||||||||||||||||||||||||||||||||||
Db 5691 GGTATTTGTGTATATTTATATAAATTGTTCTCCGT 5725

RESULT 6
AD048537/c
ID   AD048537 standard; DNA; 58326 BP.
XX
AC   AD048537;
XX
DT   12-AUG-2004 (first entry)
XX
DE   Human mannose receptor C type 2 (MRC2) genomic DNA sequence.
XX
KW   human; melanoma; single nucleotide polymorphism; SNP; neuropilin 1; NRP1;
KW   mannose receptor C type 2; MRC2; gene; ds; ENDO180.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   variation   43763
FT   /*tag= a
FT   /note= "Single nucleotide polymorphism"
XX
PN   WO200404163-A2.
XX
PD   27-MAY-2004.
XX
PF   06-NOV-2003; 2003WO-US035876.
XX
PR   06-NOV-2002; 2002US-0424475P.
PR   23-JUL-2003; 2003US-0489703P.
XX
PA   (SEQU-) SEQUENOM INC.
XX
PI   Roth RB, Nelson MR, Braun A, Kammerer SM;
XX
DR   WPI; 2004-411720/38.
XX
PT   Identifying a subject at risk of melanoma, useful for treating melanoma,
PT   comprises detecting the presence or absence of one or more polymorphic
PT   variations associated with melanoma in a nucleic acid sample from a

```
PT    subject.
XX
PS    Claim 23; SEQ ID NO 3; 176pp; English.
XX
CC    The invention comprises a method for identifying a subject at risk of
CC    melanoma. The invention involves detecting the presence or absence of one
CC    or more polymorphic variations associated with melanoma in the neuropilin
CC    1 (NRP1) or mannose receptor C type 2 (MRC2) genes. The method of the
CC    invention is useful for identifying subjects at risk and treating
CC    melanoma. The present nucleic acid represents the genomic DNA sequence
CC    for human MRC2.
XX
SQ    Sequence 58326 BP; 13666 A; 14813 C; 15614 G; 14232 T; 0 U; 1 Other;

Query Match         3.5%;      Score 196.2;  DB 12;  Length 58326;
Best Local Similarity  70.0%;  Pred. No. 5.2e-27;
Matches 278;   Conservative   0;  Mismatches 118;   Indels   1;  Gaps   1;

Qy     3262 CTTTTAAATTTTTAAAAAAATAGATGAGTGTGGTCGCTCATGCCTGTAATCCCAACACT 3321
            |||||||||| ||| || |||||||||| ||| ||||||||  ||||||||||||||||
Db    18059 CTATAAAGTTTTTTACAGGGAAAGCCAGGTGCGGTGGCTCACACCTGTAATCCCAGCACT 18000

Qy     3322 TTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAAC 3381
            ||||||||||||| || ||| ||||||| |||| ||||||||||||||||||||||||||
Db    17999 TTGGGAGGCTGAGGCAGGCAGTGGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTGGGCAAC 17940

Qy     3382 ACAGCAAGACCCCATATCTAAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTG 3441
            |||||||| |||| ||| ||||||||||||||||| |||| ||||||||||| |||||||
Db    17939 ATGGCAAAAACCCATCTCTACATAAAATACAGAAAAAAAAAATTAGCTGAGTTGGTGGTG 17880

Qy     3442 CTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAG 3501
            ||||||||||| |||| ||||||| || |||||||||||||| |||||||||||||||||
Db    17879 CACACCTGTGGTCCCTGCTTACTCAGGAGTCTCAGGAGTCTGAGATGGGAGGATTGCTTGAGCCCAGGAG 17820

Qy     3502 GTTGAGGCTGCAGTGATCCATGAAGCGCGCTTGCTACACT-CAGTCTGGGTGACAGTGCAAG 3560
            ||||||||||| |||||| ||||||||||   ||| ||||  ||| ||||||||||||||||
Db    17819 ATTGAGGCTGCAGTGAGCCAAGATCATGCCCACTGTACTCCCAGCCTGGGCAACAGAGCAAG 17760

Qy     3561 AAGCTGTCTCAAAAAAATAATAAATAAAAATAACTTTTAAAAAACAAAAATTAATTAA 3620
            |||||||||| ||| |||||| |||||||| |||| ||||||||| |||| ||| |||
Db    17759 ACTCCATCTCAAAAAAATAAAATAAAATAAAAAAATATATATATATATATATATATATATA 17700

Qy     3621 ATTTTAAAAACAACAACAACACTAGAGATGTTTGCAAAT 3657
            | |||| ||||| |   |||||  |||||||| ||
Db    17699 TATATATATACACACACATATTATGTGTGTATAT 17663
```

```
RESULT 7
ADP75180/c
ID    ADP75180 standard; DNA; 304905 BP.
XX
AC    ADP75180;
XX
DT    12-AUG-2004  (first entry)
XX
DE    Human Endophilin 2 gene.
XX
KW    Human; chromosome 19; ds; gene; ADAM19; Endophilin 1; Endophilin 2; NRG2;
KW    ADAMTS2; a disintegrin and metalloprotease; neuroregulin 2; SNP;
KW    single nucleotide polymorphism;
KW    a disintegrin and metalloprotease with thrombospondin type1 motif 2;
KW    asthma; atopy; obesity; inflammatory bowel disease; respiratory disorder.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    variation       replace(276365,T)
FT                    /*tag= a
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(276366,G)
FT                    /*tag= b
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(278194..278197,CT)
```

```
FT    variation       replace(278474,A)
FT                    /*tag= c
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(278990,C)
FT                    /*tag= d
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(279072,G)
FT                    /*tag= e
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(280049,T)
FT                    /*tag= f
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(280107,A)
FT                    /*tag= g
FT                    /standard_name= "Single nucleotide polymorphism"
FT    variation       replace(280124,T)
FT                    /*tag= h
FT                    /standard_name= "Single nucleotide polymorphism"
FT                    /*tag= i
FT                    /standard_name= "Single nucleotide polymorphism"
XX
PN    WO200303031594-A2.
XX
PD    17-APR-2003.
XX
PF    11-OCT-2002; 2002WO-US032700.
XX
PR    11-OCT-2001; 2001US-0328424P.
XX
PA    (GENO-) GENOME THERAPEUTICS CORP.
XX
PI    Keith T,    Little RD,    Van Eerdewegh P,    Dupuis J,    Del Mastro RG;
PI    Allen K;
XX
DR    WPI; 2003-381712/36.
XX
PT    New isolated nucleic acid or alternate splice variant, useful for
PT    diagnosing and treating a disintegrin and metalloprotease (ADAM) or
PT    interactor gene-associated disorder; e.g. asthma, atopy, obesity or
PT    inflammatory bowel disease.
XX
PS    Claim 2; SEQ ID NO 1; 33pp; English.
XX
CC    The invention relates to an isolated nucleic acid or alternate splice
CC    variant comprising a nucleotide sequence containing at least one of the
CC    single nucleotide polymorphisms given in the specification, a nucleotide
CC    sequence having at least 15 contiguous nucleotides of them, or
CC    complements of them. The genes are ADAM19 (a disintegrin and
CC    metalloprotease 19, also known as gene 845), NRG2 (neuroregulin 2, also
CC    known as gene 847), endophilin 1 (also known as gene 874), endophilin 2
CC    (also known as gene 803) and ADAMTS2 (a disintegrin and metalloprotease
CC    with thrombospondin type1 motif 2, also known as gene 962). Also included
CC    are a vector comprising the isolated nucleic acid (or alternate splice
CC    variant), a host cell containing the vector, an isolated polypeptide
CC    encoded by the novel nucleic acid (or alternate splice variant), an
CC    antibody or antibody fragment that binds to the polypeptide,
CC    pharmaceutical compositions (comprising the nucleic acid or alternate
CC    splice variant, vector, polypeptide or antibody, and a carrier,
CC    excipient or diluent), a kit for detecting a disintegrin and
CC    metalloprotease (ADAM) gene nucleotide sequence (comprising the isolated
CC    nucleic acid or alternate splice variant, antibody or antibody fragment,
CC    and at least one component to detect the hybridisation of the variant or
CC    the binding of the antibody to an ADAM gene amino acid sequence), a kit
CC    for detecting an interactor gene amino acid sequence (comprising the
CC    antibody or antibody fragment, and at least one component to detect the
CC    binding of the antibody to the interactor gene amino acid sequence),
CC    diagnosing an ADAM or interactor gene-associated disorder or a
CC    respiratory disorder in a human subject, determining an ADAM or
CC    interactor gene pharmacogenetic profile in a human subject, identifying
CC    an orthologue of a human ADAM or interactor gene, treating an ADAM or
CC    interactor gene-associated disorder (or a respiratory disorder) by
CC    administering the pharmaceutical composition, a transgenic mouse (whose
CC    genome comprises an introduced null mutation in an endogenous gene that
```

CC   is orthologous to a human ADAM gene), making a homozygous transgenic
CC   knockout mouse, forming a crystal of the isolated polypeptide, a cell
CC   line comprising the isolated nucleic acid or alternate splice variant, a
CC   biochip comprising the isolated nucleic acid or alternate splice variant,
CC   an isolated nucleic acid probe or primer comprising at least 8 contiguous
CC   nucleotides of the nucleic acid, an isolated antisense nucleic acid,
CC   identifying an ADAM or interactor gene ligand and an isolated nucleic
CC   acid variant of Gene 803, 845, 847, 874 or 962. The nucleic acid or
CC   alternate splice variants, methods, kits and antibody/antibody fragment
CC   are useful for diagnosing and treating an ADAM or interactor gene-
CC   associated disorder, e.g. asthma, atopy, obesity or inflammatory bowel
CC   disease. The present sequence is a gene (or gene fragment) for one of the
CC   above mentioned genes.
XX
SQ   Sequence 304905 BP; 71522 A; 82757 C; 84652 G; 65974 T; 0 U; 0 Other;

Query Match   3.5%;   Score 196;   DB 11;   Length 304905;
Best Local Similarity 73.3%;   Pred. No. 7.1e-27;
Matches 264; Conservative   0; Mismatches 95; Indels   1; Gaps   1;

Qy   3279 AAAAATAGATGAGTGTGGTGCCGGTAATGCCTGTAATCCAACACTTTGGGAGCCGGGTCGG 3338
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   156005 AATCTCGTGGGTGCCGGTAGCTCACACTGTAATCCCAGCAGCTTGGGAGGCCAAGGTGG 155946

Qy   3339 GAGGATAGCTTGAGTCCAGCAGTTTGAGACAGTCAGGGCAACACAGCAAGACCCCATAT 3398
                 |||||||| ||||||||||||||||||||||||||||||| |||||||||||||
Db   155945 GAGGATTGCTTAAGCCCAGGAGTTTGAGACCCAGGAGTTTGGGCAACATGGTGAAACCCCATGT 155886

Qy   3399 CTAAAAAAACAAAACAAAAACAAAATTACCTCGGGTATGGTTGTGCTCACCTGTAGTCCAAG 3458
                 ||||||| |||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   155885 CTACAAAAAAAATACAAAAACGTTAGCCAGGTGGTTGGTGCCCAAACCTGTAGTTTCAG 155826

Qy   3459 CTACACAGGAAGCTGAGGCAGGAGGATCACTGGGAAGGAGCTGAGGCTGAGTGAT 3518
                 |||||| |||||||||||||||||||||||||||||||||||||||||||
Db   155825 CTGCTCAGGAGGCTGAGGTGGGAGGATCACCTGAGCCCAGGAGTTGAGGCTGCAGTGAG 155766

Qy   3519 CCATGAACGGCCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAGAGCTGTCTCAAAAATA 3577
                 |||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db   155765 TCGTGATCGCGCCACTGCACTGCACCTGGGTGACAGTAAGACCCCGTCTCAAAAAAA 155706

Qy   3578 ATAAATAAATAAAATAACTTTTAAAAAACAAAAATTAATTAAAATTTTAAAAACAACAACA 3637
                 ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   155705 AAAAAAAAAACAACACAACCACGAAAACTCAATCTCTCTGTGTCTTAAAAAAAAAGCA 155646

RESULT 8
ABX56555/c
ID   ABX56555 standard; DNA; 118384 BP.
XX
AC   ABX56555;
XX
DT   20-FEB-2003 (first entry)
XX
DE   Human autoimmune disease related protein PAT1 gene region #1.
XX
KW   PAT1; human; autoimmune disease; psoriasis; type I diabetes;
KW   rheumatoid arthritis; cation-chloride transport; gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200280842-A2.
XX
PD   17-OCT-2002.
XX
PF   04-APR-2002; 2002WO-SE000669.
XX
PR   05-APR-2001; 2001GB-00008587.
PR   05-APR-2001; 2001GB-00008589.
PR   05-APR-2001; 2001GB-00008590.
XX
PA   (ESTE-) ESTETECON AB.
XX
PI   Swanbeck G, Purvis I, Dykes C, Hewett D;

XX
DR   WPI; 2003-058462/05.
XX
PT   New PAT1 genes and polypeptides, useful for diagnosing the presence of,
PT   or a susceptibility to, an autoimmune disease (e.g. psoriasis, type I
PT   diabetes or rheumatoid arthritis), or screening agents for treating the
PT   autoimmune diseases.
XX
PS   Claim 18; Page 74-104; 114pp; English.
XX
CC   The invention describes an isolated polypeptide expressed from the PAT1
CC   (gene) region, its homologue, variant or fragment. The polypeptide or
CC   polynucleotide encoding it are useful for diagnosing the presence of, or
CC   a susceptibility to, an autoimmune disease such as psoriasis, type I
CC   diabetes or rheumatoid arthritis. The polynucleotide or polypeptide are
CC   also useful for screening its modulators. The agent that modulates PAT1,
CC   or the component that regulates cation-chloride transport, is affected by
CC   action-chloride transport, regulates PAT1 or is affected by PAT1 is
CC   useful for treating, or in the manufacture of a medicament for preventing
CC   or treating, an autoimmune disease. This sequence represents a
CC   polynucleotide from the human PAT1 gene
XX
SQ   Sequence 118384 BP; 31464 A; 26095 C; 27139 G; 33686 T; 0 U; 0 Other;

Query Match   3.4%;   Score 193.4;   DB 10;   Length 118384;
Best Local Similarity 68.8%;   Pred. No. 1.9e-26;
Matches 295; Conservative   0; Mismatches 131; Indels   3; Gaps   2;

Qy   3276 AAAAAAATAGATGAGTGTGGTGGCTCATGCCTGTAATCCAACACTTTGGGAAGCCGGGT 3335
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114643 AATAAAAAGGCTGAGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAAGCCAAGG 114584

Qy   3336 CGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACAGTCAGCGGCAACACAGCAAGACCCCA 3395
                 |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   114583 CAGGTGGATTACTTGAGGTTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCTG 114524

Qy   3396 TATCTAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCC 3455
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114523 TCTCTATTAAA--AACACAAAACAAAATTAGCTGGGTGTGGTTGTGTACCGCCTGTAATCC 114466

Qy   3456 AGGCTACACAGGAGAGCTGAGGCAGGAGGATCACTGAGCCCAGGAGGTTGAGGCTGCAGT 3515
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114465 CAGCTACTTGGGAGAGCTGAGGTGAGGGAGGATCACTTGAACCCAGGAGGTGGAGGTTGTAGC 114406

Qy   3516 GATCCATGAACGACGCGTGCTACACT-CAGTCTGGGTGCGAAGAAGCTGTCTCAAAA 3574
                 |||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   114405 GAGCCAAGATCATGCCACTGCACTCCATCCTGGGTGATAGAGCAAGACTCCATCTCAAAA 114346

Qy   3575 ATAATAAATAAATAAAATAACTTTTAAAAAACAAAAATTAATTAAATTTTAAAAACACA 3634
                 |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   114345 ATGAAAAGAAAAACAAAAACAAAATAAGATAAAGTATAGTATACTAAAATATATAAA 114286

Qy   3635 ACACACTAGAGATGTTTGCAAATTGATTATTTGGGAGTCACATGTCTCATATCCCTGGAAGTTAATTT 3694
                 || ||||||||||||||||||||||||||||||||||||||||
Db   114285 CCAGTAACATGTGTTGTTTATTATCAAGTAATTATGTACTGTATATAATTGTTATTATATGT 114226

Qy   3695 AAAAATATTT 3703
                 ||
Db   114225 GCTATACTT 114217

RESULT 9
AAC74312
ID   AAC74312 standard; cDNA; 1384 BP.
XX
AC   AAC74312;
XX
DT   02-FEB-2001 (first entry)
XX
DE   Human secreted protein gene 33 SEQ ID NO:43.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;

```
Db   1116 ATGGTGATACCCCACCTCTACTAAAAAAAAAAAAAAAAAAAAAAATTAGCCAGTGTG 1175
                                                  ||||||||||||||
Qy   3436 GTTGTGTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGCAGGAGGATCACTTGAGCC 3495
                                              ||||||| ||||||| ||||||
Db   1176 GTGGCGTGCACCTGTAATCCCAGCTACTCGGAGGCTGAGGCATGATGAATCACTTGAACC 1235

Qy   3496 CAGGAGGTTGAGGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAG 3554
                ||||||                ||||| ||      || |||||  ||||||||||||
Db   1236 TGGGAGGCAGAGGTTGCAGCGAGTGAGTCAAGATTGCACCACTGCACTCCAGCCTGGGCAACAG 1295

Qy   3555 TGCAAGAAGCTGTCTCAAAAATAAATAAATAAAATAAACTTTTAAAAACAAAAATT 3614
                |||| ||||||||||| |||||||||||||| |||||||| ||||| |||||
Db   1296 AGTGAGACTCCGTCTCAAAAAAACACACACAAAAAAAAAAAAAAAAAAAAAAAA 1355

Qy   3615 AATTAAATTTTAAAAAACACACACA 3639
                || ||| ||||| |||||
Db   1356 AAAAAAAAAAAAAAAAAAAAAAA 1380
```

RESULT 10
ADQ19501
ID   ADQ19501 standard; DNA; 135005 BP.
XX
AC   ADQ19501;
XX
DT   26-AUG-2004 (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 2320.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
XX   WO200048938-A2.
PN
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N, Ginsburg WM, Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 2320; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 135005 BP; 32951 A; 33605 C; 32661 G; 35788 T; 0 U; 0 Other;

Query Match   3.3%; Score 188.4; DB 12; Length 135005;
Best Local Similarity 71.8%; Pred. No. 1.7e-25;
```

---

```
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; vulnerary; gene therapy; angiogenesis;
KW   autoimmune disease; hyperproliferative disorder; infection; skin aging;
KW   wound healing; cardiovascular disorder; cerebrovascular disorder;
KW   nervous system disorder; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO20005790 3-A2.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US007525.
XX
PR   26-MAR-1999;  99US-0126595P.
PR   22-DEC-1999;  99US-0171549P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM, Komatsoulis G;
XX
DR   WPI; 2000-594630/56.
DR   P-PSDB; AAB39342.
XX
PT   New nucleic acid molecules encoding 48 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; Page 339-340; 395pp; English.
XX
CC   The polynucleotide sequences given in AAC74280 to AAC74327 encode the
CC   human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
CC   AAB39400 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC   neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC   and vulnerary. The polynucleotides and polypeptides can be used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They can also be used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors and other nutritional
CC   components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC   the exemplification of the present invention
XX
SQ   Sequence 1384 BP; 443 A; 247 C; 254 G; 440 T; 0 U; 0 Other;

Query Match             3.3%;  Score 189.4;  DB 3;  Length 1384;
Best Local Similarity  71.9%;  Pred. No. 6e-26;
Matches  277;  Conservative  0;  Mismatches 101;  Indels  7;  Gaps  2;

Qy   3262 CTTTTAAATTTTTTAAAAAAATAGATGAGTGTGGTCGGTCATGCCTGTAATCCCAACACT 3321
                ||||  ||| |          |||||||| ||||||||||||||||
Db    996 CTTTAAAAATGTAGGAAAAATGGCTGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACT 1055

Qy   3322 TTGGGAGGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAAC 3381
                |||||||||| ||||||| || |||||||||||||| |||||||||||||
Db   1056 TTGGGAGGCCAAGGCGGCGACGGATCTCTTGAGGTCAGGAGTTTGAGACCAGCCAGCCAAC 1115

Qy   3382 ACAGCAAGACCCCATATCT------AAAAAAACAAAACAAAACAAAATTACCTGGGTATG 3435
                 | |||||||| |||    |||||||| |||  |||||||| | |||||
```

CC hyperproliferative disorders; and hyperlipidemic disorders). The present
CC DNA sequence represents a human clusterin gene sequence
XX
SQ Sequence 8133 BP; 1982 A; 2024 C; 2002 G; 2125 T; 0 U; 0 Other;

Query Match   3.3%; Score 187.6; DB 6; Length 8133;
Best Local Similarity 74.6%; Pred. No. 1.7e-25;
Matches 249; Conservative 0; Mismatches 84; Indels 1; Gaps 1;

Qy 3264 TTTAAATTTTTAAAAAAATAGATGAGTGTGGTCGCTGAATCCCAACACTTT 3323
         ||||||| |||||||||||||||||| ||||| ||||||||||||||||||
Db 1057 TTTAAAGAGGAAAAAAACTGAGCTCGGCATGCGTGGTCATGCCTGTGATGCCTCAGGGCAACTTT 1116

Qy 3324 GGGAAGCCGGGTCGGGAGGATAGCTTGAGTTCCAGCAGTTTGAGACCAGTCAGGGCAACAC 3383
         ||| ||||||| ||||||||||| |||||||||||||||||||||||||||||||||||||
Db 1117 GAGAGGCCAAGGCGGGAGGATCATTGAGGCGGGAGGTTTGAGACCGGTCAGGCCTGGCCAAGAT 1176

Qy 3384 AGCAAGACCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTATGGTTGTGCT 3443
         |||||| |||||| ||| ||||||||||||||||| |||||||||||||||||||
Db 1177 AGTGAGACCCTGTCTCTCTACAAAATAAAAACTTAAAAAATTAGCCGGGTGGTGGTGCA 1236

Qy 3444 CACCTGTAGTCCAAGCTACACAGGAGCTGAGGCAGAAGGATCACTGAGGCCAGGAGGT 3503
         |||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||
Db 1237 CACCTGTGTCTCAGCTATCTCGGGAGGCTACTCCGGGAGGCCAAGGCAAGAGAGTCACCTGAGCCTGGGAAGTT 1296

Qy 3504 TGAGGCTGCAGTGATCCATGAACGCGCTGCTCACACT--CAGTCTGGGTGACAGTGCAAGAA 3562
         ||| ||||||||||| |||||||||||||||||||| |||||||||||||||| |||||||
Db 1297 GGAGGCTGCAGTGAGCTATGATTGCACCATTGCACCATTCCACGCTGGGCAACAGAGTGAGAC 1356

Qy 3563 GCTGTCTCAAAAATAATAATAAATAATAAAAATAATTAAAAAATAAC 3596
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 1357 CCTGTCTCTAAATTAAAAAATAAAATAATAATAAAAATAAC 1390

RESULT 12
ADJ12659
ID ADJ12659 standard; DNA; 18861 BP.
XX
AC ADJ12659;
XX
DT 20-MAY-2004 (first entry)
XX
DE DNA fragment of a BAC clone that encodes a human secreted protein Seq513.
XX
KW human; secreted; cancer; haematopoietic disease; anaemia;
KW multiple myeloma; reproductive system disorder; prostatitis;
KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;
KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;
KW fetal alcohol syndrome; Down's syndrome; excretory disease;
KW urinary incontinence; renal disorder; neural; sensory disease;
KW Alzheimer's disease; meningitis; respiratory disease; emphysema;
KW occupational lung disease; endocrine disease; diabetes;
KW glomerulonephritis; digestive disease; portal hypertension;
KW irritable bowel syndrome; epithelial disease; scleroderma;
KW epidermolysis bullosa; cytostatic; antianemic; antiarthritic;
KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
KW antipsoriatic; antibacterial; osteopathic; dermatological; antigout;
KW immunomodulator; antiarrhythmic; cardiant; nootropic; antilipemic;
KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
KW antidiabetic; anabolic; hypertensive; vulnerary; ds.
XX
OS Homo sapiens.
XX
PN US2004010132-A1.
XX
PD 15-JAN-2004.
XX
PF 30-OCT-2001; 2001US-00984429.
XX
PR 09-OCT-1997; 97US-0061463P.
PR 09-OCT-1997; 97US-0061527P.
PR 09-OCT-1997; 97US-0061529P.
PR 09-OCT-1997; 97US-0061532P.

Matches 260; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

Qy 3263 TTTAAATTTTTAAAAAAATAGATGAGTGTGGTCGCTGAATCCCAACACTT 3322
         |||||| ||||||||||||||||||||||||| |||||||||||||||||
Db 26435 TTTTATAAAAATGGTGACAGTGGCTGAGTGCAGTCCTGTAATCCCAACACTT 26494

Qy 3323 TGGGAAGCCGGGTCGGGAGGATAGCTTGAGTTCCAGCAGTTCAGGGCAACA 3382
         |||||||||||| |||||||||| |||||||| ||| |||||||||||||||
Db 26495 TGGGAGGCTGAGGCAGGGGGATTGCTTGAGCCCATGAGTTTGAGACCAGCTGGCCAACA 26554

Qy 3383 CAGCAAGACCCCATATCTAAAAAAACAAAAACAAAATTACCTGGGTATGGTTGTGC 3442
         |||||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||
Db 26555 GGGTAAAATCCCATCTCAACAAAAATACACACAAAAATAGTTGGGCATGGTGGTGT 26614

Qy 3443 TCACCTGTAGTCCAGCTACACAGGAGCTGAGGCAGGAGGATCACTGAGCCCAGGAGG 3502
         | ||||||||||||||| ||| |||||||| |||||||||||||||||||||||||||
Db 26615 ACATCTGTGTCCCAGCTACTCAGGAGGCCGAGGTTGGGAGGACCAGCCGAGGCCAGGAGG 26674

Qy 3503 TTGAGGCTGCAGTGATCCATGAACCGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGA 3561
         | |||| |||||||||||||||| ||||||||||||| ||||||||||| ||||||||||
Db 26675 TCTAGACTGCAGCGAGCTGTGATTGTGCCCACTGCACTCCCAGTCTCGGCAACAGAGTGAGA 26734

Qy 3562 AGCTGTCTCAAAAATAATAATAAATAATAAAAATAATTAAAAAA 3621
         |||||||||||||||||||| ||||||||||| |||||||||||
Db 26735 CCCTGTCTCTAAAATAATAATAATAATAATAAAAATAATTAAAATTAAAATTAAA 26794

Qy 3622 TT 3623
         ||
Db 26795 AT 26796

RESULT 11
ABN99663
ID ABN99663 standard; DNA; 8133 BP.
XX
AC ABN99663;
XX
DT 16-AUG-2002 (first entry)
XX
DE Human clusterin gene sequence 2.
XX
KW Human; antisense inhibition; antisense oligonucleotide; clusterin;
KW hypercholesterolaemia; cardiovascular disorder; ds;
KW hyperproliferative disorder; hyperlipidemic disorder.
XX
OS Homo sapiens.
XX
PN WO200222635-A1.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US028235.
XX
PR 11-SEP-2000; 2000US-00659791.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Monia BP, Freier SM;
XX
DR WPI; 2002-404805/43.
XX
PS Example 15; Page 93-98; 125pp; English.
XX
PT Novel antisense compound targeted to nucleic acid molecule encoding
PT clusterin, useful for treating animal having disease associated with
PT clusterin such as hyperlipidemic disorder, cardiovascular disorder.
XX
CC The invention comprises antisense oligonucleotides that are capable of
CC inhibiting expression of the human clusterin gene. The antisense
CC oligonucleotides of the invention are useful for inhibiting the
CC expression of clusterin in cells. The antisense oligonucleotides are also
CC useful for treating an animal with a disease or condition associated with
CC clusterin (e.g. hypercholesterolaemia; cardiovascular disorders;

```
Db  5617  |||||||||  ||  ||||||  |||||||  ||  ||  || ||||||||||||||  ||||||
          GGAGGCCGAGGTGGGAGAATCACTTGAGCTCAGGAGTTTGAGAGTTTGAGCGAGCAACATA  5676

Qy  3385  GCAAGACCCCATATCTAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGCTC  3444
          |||| ||| |||||  ||||  |||||||||||||||||||||||  ||||  ||| ||||
Db  5677  GCAAGACTTCATCTCTCACTAAAAGAAGAAGTTAAAAAATTAGCCAGGCGTGGTGGCTCAC  5736

Qy  3445  ACCTGTAGTCCAAGCTACACAGGAGGCTGAGGCAGGAGGCAGGAGGCTGAGGCCCAGGAGGTT  3504
          ||||||||||||  ||||||| ||||||||||||||||||||||||  ||||||||||||
Db  5737  ACCTGTAGTCCCAGCTACTCAGAAAGCTGAGGCAGGAGGATCACTTGAGTCTGGGAGGTT  5796

Qy  3505  GAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAG  3563
          |||||||||||||||||||||  ||| ||| ||||  |||||||||  || |||||||||
Db  5797  GAGGGTGCAGTGAGCCATGATCGCACCACTGCACTCCACTCCAGTCTGGGTGACAGAGTACGACT  5856

Qy  3564  CTGTCTCAAAAATAATAA  3581
          |||||||||  |||| ||
Db  5857  CTGTCTCCAAAAGACAAA  5874
```

```
RESULT 13
AAD52898
ID   AAD52898 standard; DNA; 47999 BP.
XX
AC   AAD52898;
XX
DT   14-MAY-2003  (first entry)
XX
DE   Human tweety homologue 2 (TTYH2) gene.
XX
KW   Human; tweety homologue 2; TTYH2; therapy; cancer; tumour; cytostatic;
KW   diagnostic marker; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1936..2074
FT                   /tag= b
FT                   /number= 1
FT   CDS             1946..45753
FT                   /tag= a
FT                   /product= "Human TTYH2 protein"
FT   intron          2075..10376
FT                   /tag= c
FT   exon            10377..10549
FT                   /tag= d
FT                   /number= 2
FT   intron          10550..16622
FT                   /tag= e
FT   exon            16623..16734
FT                   /tag= f
FT                   /number= 3
FT   intron          16735..23223
FT                   /tag= g
FT   exon            23224..23444
FT                   /tag= h
FT                   /number= 4
FT   intron          23445..28299
FT                   /tag= i
FT   exon            28300..28395
FT                   /tag= j
FT                   /number= 5
FT   intron          28396..28902
FT                   /tag= k
FT   exon            28903..28975
FT                   /tag= 1
FT                   /number= 6
FT   intron          28976..35372
FT                   /tag= m
FT   exon            35373..35442
FT                   /tag= n
FT                   /number= 7
FT   intron          35443..35705
```

PR   09-OCT-1997;   97US-0061536P.
PR   09-OCT-1997;   97US-0071498P.
PR   08-OCT-1998;   98WO-US021142.
PR   08-APR-1999;   99US-00288143.
PR   01-NOV-2000;   2000US-0244591P.
XX
PA   (ROSE/) ROSEN C A.
PA   (BREW/) BREWER L A.
PA   (DUAN/) DUAN R D.
PA   (RUBE/) RUBEN S M.
PA   (FLOR/) FLORENCE K A.
PA   (GREE/) GREENE J M.
PA   (YOUN/) YOUNG P E.
PA   (FERR/) FERRIE A M.
PA   (YUGG/) YU G.
PA   (FLOR/) FLORENCE C.
PA   (EBNE/) EBNER R.
PA   (OLSE/) OLSEN H.
XX
PI   Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM;
PI   Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;
XX
DR   WPI; 2004-090518/09.
XX
PT   New isolated nucleic acids and polypeptides, useful for diagnosing,
PT   treating, preventing or ameliorating diseases or disorders e.g. cancer,
PT   anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's
PT   disease.
XX
PS   Disclosure; SEQ ID NO 513; 286pp; English.
XX
CC   This invention relates to novel polynucleotides encoding human secreted
CC   proteins. Specifically, it refers to the vectors, host cells, recombinant
CC   and synthetic methods for producing human polynucleotides, polypeptides
CC   and antibodies. Furthermore, it relates to screening methods to identify
CC   agonists and antagonists that can be used to inhibit or enhance the
CC   production and function of the secreted proteins. The present invention
CC   describes these compositions as useful for diagnosing, treating or
CC   preventing disorders such as cancer, haematopoietic diseases including
CC   anaemia and multiple myeloma, reproductive system disorders including
CC   prostatitis and inguinal hernia, musculoskeletal diseases including
CC   systemic lupus erythematosus and gout, cardiovascular disease including
CC   arrhythmia and hypernatraemia, mixed fetal diseases including fetal
CC   alcohol syndrome and Down's syndrome, excretory diseases including
CC   urinary incontinence and renal disorders, neural or sensory disease
CC   including Alzheimer's disease and meningitis, respiratory disease
CC   including emphysema and occupational lung disease, endocrine diseases
CC   including diabetes and glomerulonephritis, digestive diseases including
CC   portal hypertension and irritable bowel syndrome and connective tissue or
CC   epithelial diseases including scleroderma and epidermolysis bullosa. As
CC   such, there are various activites such as cytostatic, antianemic,
CC   antiarthritic, antiasthmatic, anti-HIV, immunosuppressive,
CC   antiinflammatory, antipsoriatic, antibacterial, osteopathic,
CC   dermatological, antigout, immunomodulator, antiarrhythmic, cardiant,
CC   nootropic, antilipemic, nephrotropic, uropathic, neuroprotective,
CC   antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and
CC   vulnerary. This polynucleotide is a DNA fragment of a BAC clone that
CC   encodes a human secreted protein of the invention. NOTE: This sequence
CC   does not appear in the printed specification but has been obtained in
CC   electronic format from the US patent office at the following web site
CC   www.seqdata.uspto.gov/sequence.html; Document ID: 2004010133.
XX

```
SQ   Sequence 18861 BP; 4607 A; 4729 C; 4409 G; 5116 T; 0 U; 0 Other;

Query Match        3.3%;   Score 186;  DB 12;  Length 18861;
Best Local Similarity   76.1%;  Pred. No. 3.7e-25;
Matches  242;  Conservative   0;  Mismatches  75;  Indels  1;  Gaps  1;

Qy  3265  TTAAATTTTTAAAAAAAATAGATGAGTGTGGCTCATGCCTGTAATCCAACACTTTG  3324
          ||   |||  ||||||||  |||||||| ||||||  ||||||||||||||  ||||||
Db  5557  TAAAAGTTGAAAAAAAAAAATGACTAGGTGCAGCGGCTCTAATCCTCCAGCGCTTTG  5616

Qy  3325  GGAAGCCGGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACA  3384
```

```
FT  exon                 /*tag= o
FT                 35706. .35761
FT                       /*tag= p
FT                       /number= 8
FT  intron         35762. .36266
FT                       /*tag= q
FT  exon           36267. .36359
FT                       /*tag= r
FT                       /number= 9
FT  intron         36360. .36591
FT                       /*tag= s
FT  exon           36592. .36684
FT                       /*tag= t
FT                       /number= 10
FT  intron         36685. .38529
FT                       /*tag= u
FT  exon           38530. .38672
FT                       /*tag= v
FT                       /number= 11
FT  intron         38673. .39376
FT                       /*tag= w
FT  exon           39377. .39562
FT                       /*tag= x
FT                       /number= 12
FT  intron         39563. .40050
FT                       /*tag= y
FT  exon           40051. .40129
FT                       /*tag= z
FT                       /number= 13
FT  intron         40130. .45672
FT                       /*tag= aa
FT  exon           45673. .47158
FT                       /*tag= ab
FT                       /number= 14
XX
PN  WO20029629-A1.
XX
PD  21-NOV-2002.
XX
PF  14-MAY-2002; 2002WO-AU000591.
XX
PR  14-MAY-2001; 2001AU-00004971.
XX
PA  (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX
PI  Clements JA;
XX
DR  WPI; 2003-129264/12.
DR  P-PSDB; AAE34613.
XX
PT  New human tweety homolog 2 polypeptides and polynucleotides, useful for
PT  producing an antigen-binding molecule that is immuno-interactive with the
PT  polypeptide or as diagnostic markers for cancers.
XX
PS  Claim 10; Page 128-156; 176pp; English.
XX
CC  The invention relates to human tweety homologue 2 (TTYH2) polypeptide and
CC  polynucleotide sequence. TTYH2 is useful for producing an antigen-binding
CC  molecule that is immuno-interactive with the polypeptide. The agent is
CC  useful for manufacturing a medicament for restoring a normal level and/or
CC  functional activity of TTYH2 expression in a patient, and for treating or
CC  preventing cancer or tumour. TTYH2 sequences may also be used to provide
CC  both drug targets and regulators to promote or inhibit one or more
CC  activities, and to provide diagnostic markers for cancers. The present
CC  sequence is human TTYH2 gene
XX
SQ  Sequence 47999 BP; 11178 A; 12727 C; 12874 G; 11209 T; 0 U; 11 Other;

Query Match       3.3%;   Score 185.4; DB 8; Length 47999;
Best Local Similarity   74.9%;   Pred. No. 5.5e-25;
Matches 245;  Conservative 0; Mismatches 81; Indels 1; Gaps 1;

Qy  3284  AGATGAGTGGTGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAGGCCGGGTCGGGAGGA 3343
```

```
Db  18395  AGTTGAGTGTGGTGGCTCACGCCTGAAATCCAGCACTTTGGGAGGCCAAGGTGGGAGAA 18454
           ||||||||| ||||||||||   ||  |||| ||||||||||   ||||||||||||||
Qy  3344   TAGCTTGAGTCCAGCAGTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTAA 3403

Db  18455  TTGCCTGAGGCCAGAGTTCAAGACCAGCCTGGCAACATAGCAAGACCCTATCTCTACA 18514
           || |||||||||||| |||||  |||||| |||| ||  | ||||||||||||| ||
Qy  3404   AAACAAACAAACAAACAAATTACCTGGGTATGGTTGTGTCTCACCTGTAGTCCAAGCTACA 3463

Db  18515  AAAATTAAAAATAAAAGTAGCTGAGTGTGGTGGCACGTGTCCAACCACT 18574
           |||||||||||||||
Qy  3464   CAGGAAGCTGAGGCAGGAGGATCACTTGACCCTCAGGAGGTTGAGGCTGCAGTGATCCATG 3523

Db  18575  CAGGAGGCTGAGGTGGGAGGATTGCTTGAGCCCCAGGAGTTTGAGGCTGCAGTGAGCTATG 18634
           ||||| |||||| ||||||||| | |||| |||||||| ||||||||||||||| ||||
Qy  3524   AACGGCCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAA 3582

Db  18635  ATCACAGCACTGCCGCTCCCAGCTTGGGTGACAGAGTGAGACCTTGTCTCTAAAAAAAAAG 18694
           || ||  ||   || |||| ||||  | ||||||  |
Qy  3583   TAAATAAAAATAACTTTTAAAAACAA 3609

Db  18695  AAAAGAAAAAAAAAAGAAATGATAA 18721
           ||||| ||||||||| |
```

RESULT 14
ADL13719/c
ID   ADL13719 standard; DNA; 199878 BP.
XX
AC   ADL13719;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #251.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA, Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 251; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:

```
CC    The sequence data for this patent did not form part of the printed
CC    specification but was obtained in electronic format directly from WIPO at
CC    ftp.wipo.int/pub/published_pct_sequences).
XX
SQ    Sequence 199878 BP; 58961 A;  40585 C; 39746 G; 60450 T; 0 U; 136 Other;

Query Match           3.3%;    Score 184.8;  DB 10;  Length 199878;
Best Local Similarity  62.9%;  Pred. No. 8.5e-25;
Matches  302; Conservative   0; Mismatches  177;  Indels   1;  Gaps   1;

Qy    3143 TTTGTTGCCATTAATTTCTGTTCCTCTTAGCTTAGTAACTTTAGGATTTTTAAATAACA 3202
                |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  159286 TTTGGTTACCTTAAATAACCAAACAACCAAACTGATAATTTGGGAGATATTTTCATTAAC 159227

Qy    3203 ACTATTGAAATCATGACATACGTTTAAATGATATTATTTAAATACGTTAGGCTATAAACC 3262
                ||||       |||||||||  ||||||    ||||
Db  159226 TCTGATTTCCTGGGGCCTGACTGGCTAACTTATGCATTATTTAAATAAAACAGCAC 159167

Qy    3263 TTTTAAATTTTTAAAAAATAGATGAGTGTGGTGGCCTCATGCCTGTAATCCCAACACTT 3322
                |||       ||||||||||||||  ||| ||||||| ||| |||| |||||
Db  159166 TTTTCATCAGAAGAGAGACGGGGCTGGGTGGGTGGTAGCTCATTCCTATAAACCCAGCATTT 159107

Qy    3323 TGGGAAGCCCGGGTCGGGAGGATAGTTGAGTCCAGCAGTTTGAGACCAGTTCAGGGCAACA 3382
                ||||||       ||||||| |||||| |||||||||||||  |||||||||
Db  159106 TGGGACGCTAAGGCGGGAGGATCACTTGAGGCCCAGGGGTTCGAGACCAGCCTGGGCAACA 159047

Qy    3383 CAGCAAGACCCCATATCTAAAAAACAAAACAAAATTACCTGGGTATGGTTGTGC 3442
                ||||||        ||||||||||||||||  |  |||||||  |||| ||| ||||
Db  159046 TGGCAAGACCCCCCTCTCTACAAAAAAAATTTAAAAAAAAATTAGCCAGGTGGGGTGGTTC 158987

Qy    3443 TCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAAGGATCACTTGAGCCCAGGAGG 3502
                ||||||||  ||||||| |||||  |||||| |||||||||||||||||||||||||||||
Db  158986 ATGTTTGTAGTCCCCAGCTACTCAGGAGGCTGAGGTGGGGGATCACTTGAGCCCAGGAGT 158927

Qy    3503 TTGAGGCTGCAGTGATCCATGAACCGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGA 3561
                ||||||||||||||||||||| |||||||||||||||  ||||||||||  ||||||||||
Db  158926 TTAAGGCTGCAGTGAACTATGATCATCACCCCTCTGCACTCCAGCCTGGGTGACAGAGTAAGA 158867

Qy    3562 AGCTGTGTCTCAAAAATAATAAATAATAAAAATAAACTTTAAAAATCAAAAATTATTAAA 3621
                ||||        ||||||||  ||||||   |||||||||||  ||||| ||||||||||
Db  158866 CCCCATTTCTAAGAGAAAAAAAAAAGAATTGAAGTTAAAAATTTTAAAATATTAAAGAAGGGAAA 158807
```

```
RESULT 15
ADC24703/c
ID    ADC24703 standard; DNA; 335199 BP.
XX
AC    ADC24703;
XX
DT    18-DEC-2003  (first entry)
XX
DE    Human wild type HNL4Y genomic sequence.
XX
KW    ds; nootropic; neuroleptic; tranquillizer; gene therapy; synaptogenesis;
KW    mutation; neurological disease; mental disorder; psychiatric illness;
KW    autism; Asperger syndrome; schizophrenia;
KW    attention deficit hyperactivity disorder.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT
FT    exon            10156..10298
FT                    /*tag= a
FT    intron          10299..108499
FT                    /*tag= b
FT                    /*number= 1b
FT    exon            108500..109001
FT                    /*tag= c
FT                    /*number= 2
FT    intron          109002..205868
FT                    /*tag= d
FT                    /*number= 2
```

```
FT    exon                        205869..205928
FT                                /*tag= e
FT                                /*number= 2b
FT    intron                      205929..209525
FT                                /*tag= f
FT                                /*number= 2b
FT    exon                        209526..209679
FT                                /*tag= g
FT                                /*number= 3
FT    intron                      209680..235027
FT                                /*tag= h
FT                                /*number= 3
FT    exon                        235028..235139
FT                                /*tag= i
FT                                /*number= 3b
FT    intron                      235140..238089
FT                                /*tag= j
FT                                /*number= 3b
FT    exon                        238090..238212
FT                                /*tag= k
FT                                /*number= 3t
FT    intron                      238213..310596
FT                                /*tag= l
FT                                /*number= 3t
FT    exon                        310597..310783
FT                                /*tag= m
FT                                /*number= 4
FT    intron                      310784..316138
FT                                /*tag= n
FT                                /*number= 4
FT    exon                        316139..316929
FT                                /*tag= o
FT                                /*number= 5
FT    intron                      316930..326821
FT                                /*tag= p
FT                                /*number= 5
FT    exon                        326822..330136
FT                                /*tag= q
FT                                /*number= 6
XX
PN    WO200304598-A2.
XX
PD    05-JUN-2003.
XX
PF    02-DEC-2002; 2002WO-FR004134.
XX
PR    30-NOV-2001; 2001CA-02364106.
XX
PA    (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA    (INSP ) INST PASTEUR.
PA    (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
XX
PI    Bourgeron T, Jamain S, Quach H, Betancur C, Leboyer M;
PI    Gillberg C;
XX
DR    WPI; 2003-493399/46.
XX
PT    New nucleic acid encoding mutant protein involved in synaptogenesis,
PT    useful for treatment and diagnosis of e.g. autism, Asperger syndrome, and
PT    schizophrenia.
XX
PS    Claim 10; SEQ ID NO 4; 416pp; French.
XX
CC    The invention relates to an isolated or purified polynucleotide encoding
CC    a polypeptide (the wild-type form of which is involved in synaptogenesis)
CC    that includes (at least one mutation associated with development of
CC    neurological disease and/or a predisposition to development of mental
CC    disorders or psychiatric illness. The polypeptide are used to screen for
CC    agents that modulate their activity. Also nucleic acid, polypeptide,
CC    polypeptide-specific antibodies, vectors containing he nucleic acid and
CC    host cells containing the vector, are useful as pharmaceuticals for
CC    treating mental and neurological disorders, specifically autism, Asperger
CC    syndrome, schizophrenia and attention deficit hyperactivity disorder. The
```

CC   wild-type forms of the nucleic acid and polypeptide can be used
CC   similarly. Also detecting mutations in the nucleic acid and polypeptide,
CC   or measuring activity of the polypeptide, can be used to detect
CC   biochemical disorders that affect formation of synapses and to diagnose
CC   mental disease. This sequence corresponds to the genomic sequence of the
CC   human wild type HNL4Y gene.
XX
SQ   Sequence 335199 BP; 98958 A; 63589 C; 65677 G; 106975 T; 0 U; 0 Other;

Query Match   3.2%; Score 183; DB 10; Length 335199;
Best Local Similarity   65.2%; Pred. No. 2e-24;
Matches 303; Conservative   0; Mismatches 155; Indels   7; Gaps   2;

Qy   3158 TTTCTGTTCCTCTTAGCTTAGTAACTTTAGGATTTTAAATAACAACTATTGAAATCATG 3217
            |||| || | | |||| ||| || | | | |||||
Db   239974 TTTATATATCTTATATAAAGTCATCTTTTATATATATAAAATATATTATATATTATAATT 239915

Qy   3218 ACATACGTTTAAATGATATTATTTAAATACGTTAGGCTATAAACCTTTTAAATTTTTTAA 3277
            | | | || | | | || ||| | || || | ||
Db   239914 ATGTTACAAAAATGTAATATATATATATAATATAAAATCATTACAGTGACATTGAAGAGAAAAA 239855

Qy   3278 AAAAATAGATGAGTGGTGGTCGCCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCG 3337
            |||| | | |||||||||||||| || || |||||||| || ||| | |
Db   239854 AAAATTGGCCAGGTGCGGGTGTCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTG 239795

Qy   3338 GGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGCAACACAGCAAGACCCCATA 3397
            ||||| | || ||||| ||||||||||||||| | ||| | |||| || |||
Db   239794 GGTGGATTGCCTGAGCTCAGGAGTTTGAGACCAGCCTGGGCAACACAGTAAAACCCCATC 239735

Qy   3398 TCTAAAAAACAAAAACAAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAA 3457
            ||| | | | | | | |||| |||||||| || |||| |||| |||||||||
Db   239734 TCTATGGAAATACAA------AAATTAGTTGAGCGTGGTGGTGCCACACCTGTAGTCCCA 239681

Qy   3458 GCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGTTGAGGCTGCAGTGA 3517
            |||| | |||||| || | || | | || ||| |||||| | | |||||||||
Db   239680 GCTACTGGGGAGGCTGAGACAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTTGCAGTGA 239621

Qy   3518 TCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAAT 3576
             ||||| | || ||| | ||| |||||| ||||||||| | |||||| |||||||||||
Db   239620 GCTGAGATCACGCCATTGCACTCCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAA 239561

Qy   3577 AATAAATAAATAAAAATAAACTTTTAAAAAACAAAAATTAATTAAA 3621
            |||||| |||| | ||| ||||| ||| | || ||
Db   239560 ATAAAATAAAATAAAATAGAAAAAAGAAAAAAGAAAAAGAAACCAAAA 239516

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.6
        Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:        June 23, 2005, 00:25:12 ; Search time 845 Seconds
                                (without alignments)
                            10991.132 Million cell updates/sec

Title:         US-10-030-294-1
Perfect score: 5676
Sequence:      1 cagctgttcaggatgtctg.........tttatataattgttctccgt 5676

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents NA:*
             1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
             2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
             3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
             4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
             5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
             6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1987.4	35.0	255679	4	US-09-949-016-17189	Sequence 17189, A
2	190.6	3.4	265038	4	US-09-949-016-15779	Sequence 15779, A
3	187.6	3.3	8133	3	US-09-659-791A-10	Sequence 10, Appl
4	186.4	3.3	601	4	US-09-949-016-142852	Sequence 142852,
5	186	3.3	17348	4	US-09-949-016-17403	Sequence 17403, A
6	185.6	3.3	201827	4	US-09-949-016-201827	Sequence 201827,
7	185.4	3.3	37292	4	US-09-949-016-15382	Sequence 15382, A
8	184.8	3.3	72677	4	US-09-949-016-13444	Sequence 13444, A
9	184.4	3.3	33272	4	US-09-949-016-16949	Sequence 16949, A
10	184.4	3.2	601	4	US-09-949-016-58699	Sequence 58699, A
11	184	3.2	601	4	US-09-949-016-132156	Sequence 132156,
12	183.4	3.2	601	4	US-09-949-016-119415	Sequence 119415,
13	182.8	3.2	36228	4	US-09-949-016-12256	Sequence 12256, A
14	182.8	3.2	36228	4	US-09-949-016-15468	Sequence 15468, A
15	182.6	3.2	134140	4	US-09-949-016-12672	Sequence 12672, A
16	182.6	3.2	134241	4	US-09-949-016-12424	Sequence 12424, A
17	182.6	3.2	134242	4	US-09-949-016-15813	Sequence 15813, A
18	182.6	3.2	134242	4	US-09-949-016-15814	Sequence 15814, A
19	182.6	3.2	134242	4	US-09-949-016-15815	Sequence 15815, A
20	182.4	3.2	601	4	US-09-949-016-132157	Sequence 132157,
21	182.2	3.2	36651	3	US-09-738-894A-3	Sequence 3, Appli
22	182.2	3.2	36651	4	US-09-964-469-3	Sequence 3, Appli
23	182.2	3.2	51403	4	US-09-949-016-15057	Sequence 15057, A
24	181.8	3.2	4732	4	US-09-949-016-14962	Sequence 14962, A
25	181	3.2	601	4	US-09-949-016-127367	Sequence 127367,
26	181	3.2	69924	4	US-09-949-016-15367	Sequence 15367, A
27	180.8	3.2	21920	4	US-09-949-016-15609	Sequence 15609, A
c 28	180.2	3.2	86273	4	US-09-949-016-15273	Sequence 15273, A
29	180.2	3.2	87205	4	US-09-949-016-13430	Sequence 13430, A
30	180	3.2	78720	4	US-09-949-016-12710	Sequence 12710, A
31	180	3.2	78720	4	US-09-949-016-17283	Sequence 17283, A
32	179.6	3.2	103377	4	US-09-949-016-14089	Sequence 14089, A
33	179.6	3.2	119032	4	US-09-949-016-12160	Sequence 12160, A
34	179.6	3.2	119032	4	US-09-949-016-17268	Sequence 17268, A
35	179.6	3.2	163181	4	US-09-949-016-13730	Sequence 13730, A
36	179.4	3.2	601	4	US-09-949-016-58698	Sequence 58698, A
c 37	179.2	3.2	77851	4	US-09-949-016-12508	Sequence 12508, A
38	179.2	3.2	77867	4	US-09-949-016-13211	Sequence 13211, A
c 39	179.2	3.2	77867	4	US-09-949-016-13212	Sequence 13212, A
c 40	179.2	3.2	77940	4	US-09-949-016-12509	Sequence 12509, A
c 41	179	3.2	65848	4	US-09-949-016-13285	Sequence 13285, A
c 42	178.8	3.2	152132	4	US-09-949-016-13845	Sequence 13845, A
c 43	178.8	3.1	152145	4	US-09-949-016-12371	Sequence 12371, A
c 44	178.6	3.1	186734	4	US-09-949-016-14870	Sequence 14870, A
c 45	178.6	3.1	193689	4	US-09-949-016-12350	Sequence 12350, A

```
                                ALIGNMENTS

RESULT 1
US-09-949-016-17189
; Sequence 17189, Application US/0994916
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
US-09-949-016-17189
; SEQ ID NO 17189
; LENGTH: 255679
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(255679)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17189

Query Match        35.0%;  Score 1987.4;  DB 4;  Length 255679;
Best Local Similarity  99.9%;  Pred. No. 0;
Matches 1999;  Conservative 0;  Mismatches 1;  Indels 1;  Gaps 1;

Qy    3676  TATCCCTGGAAGTTAATTAAATATTTAGAAGAGTTCTTCCTCATTCCTAGAGACGTC 3735
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  TATCCCTGGAAGTTAATTAAATATTTAGAAGAGTTCTTCCTCATTCCTAGAGACGTC   60

Qy    3736  GAATTGTAAATATCAGAGCTAGAGGAACACTAGGGCTCGCCACTCCAAAGTGTGGTCCA 3795
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  GAATTGTAAATATCAGAGCTAGAGGAACACTAGGGCTCGCCACTCCAAAGTGTGGTCCA  120

Qy    3796  AGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAGAGTCTTAGGCCTCAC 3855
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  AGGACCAGCAGCATCAAGTAACCTGGGAACGTGTTAGAAATGCAGAGTCTTAGGCCTCAC  180

Qy    3856  CCCAGACCTACTGAACCAGAATCTGCATTAACAGATTTCTAGGTGCCTCACGGCACAT 3915
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  CCCAGACCTACTGAACCAGAATCTGCATTAACAGATTTCTAGGTGCCTCACGGCACAT  240

Qy    3916  TAAAACTTGAGAAGCTCTGCACTAGAGAAATCTTCACTCCACCTTTCATTATAAATGGAATC 3975
```

```
Db    241 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TAAAACTTGAGAAGCTCTGCACTAGAAATCTTCACTCCACCTTTCATTATAAATGGAATC  300

Qy   3976 ACTTGGGCTGCTGGTCACAGGAAATGATTATTTTTAATTTCAGAACCTTCTATTTAGGTC 4035
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACTT-GGCTGCTGGTCACAGGAAATGATTATTTTTAATTTCAGAACCTTCTATTTAGGTC  359

Qy   4036 ATCTATATTTGCTAATAGCAGGGAGAAAGCCAAACTCTTTAACTGCAATTAACAAATCT  4095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 ATCTATATTTGCTAATAGCAGGGAGAAAGCCAAACTCTTTAACTGCAATTAACAAATCT   419

Qy   4096 ATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTGTGGAGCAAGCTGTTT  4155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 ATAATTAATTAGTTAAGCAATCTTCCCTTTAAGTTTTACATTTGTGGAGCAAGCTGTTT   479

Qy   4156 GATTTGGCTCAGGCCGGCCCTGTTTGTGAATTTCACAATTCACAGATGTTAGCCG      4215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 GATTTGGCTCAGGCCGGCCCTGTTTGTGAATTTCACAATTCACAGATGTTAGCCG       539

Qy   4216 CTCTCCGGGCTAAGTAAAGAAGACAGAAATGTCAAGTTTAAATAGCTTCTCCCTTCCATCCT 4275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 CTCTCGGGGCTAAGTAAAGAAGGAGAGAGAGTGTCAAGTTTAAATAGCTTCTCCCTTCCATCCT 599

Qy   4276 GGCTGAAGCAACAAATAAAATATTTTATGAAAACACATTTGAGTTAGAATTTACTTACTACAG 4335
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    600 GGCTGAAGCAACAAATAAAATATTTTTATGAAAACACATTTGAGTTAGAATTTACTTACTACAG 659

Qy   4336 GGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCCACAACTTTGACATCTCACTGA 4395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 GGAAATGTCAAATTTCTCTGAAAGGGCTTTAGATTGTCTCCACAACTTTGACATCTCACTGA 719

Qy   4396 TGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGAGATGTGAACTCACC  4455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 TGTCACCTATTTACAGGTGTGTCCTGTGACTAGGGGTGAAGGAGATGTGAACTCACC   779

Qy   4456 ATGTTAGTGACCGTTAGATCACAGAGTGTGTTTTTTCCCCGTTGTGGAGTCTCATCCTA  4515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 ATGTTAGTGACCGTTAGATCACAGAGTGTTTTTTTCCCCGTTGTGGAGTCTCATCCTA   839

Qy   4516 ACTGAGCTTCTGAATCATATTTCATTCAATTCAATTTCCAAATCCACAAAACCAGGATAAGTTTA 4575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    840 ACTGAGCTTCTGAATCATATTTCATTCAATTCAATTTCCAAATCCACAAAACCAGGATAAGTTTA 899

Qy   4576 CAGCCCATATTCAGAAAGGAAATAAATAAATTATTTTGTGTGTAGACTTTCCTGATATTACACT 4635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    900 CAGCCCATATTCAGAAAGGAAATAAATAAATTATTTTGTGTGTAGACTTTCCTGATATTACACT 959

Qy   4636 GATTTGGGAATATATGAACAATTTTATGGTTTCCTTTCGAAGTAGTCAAGTCAAAGCAA  4695
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    960 GATTTGGGAATATATGAACAATTTTTATGGTTTCCTTTCGAAGTAGTCAAGTCAAAGCAA 1019

Qy   4696 AACCAAAACAGCAAAAAACTGTAAGACATAAAGAATAGAGTGGAGCCGACTGAGAGATTA 4755
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1020 AACCAAAACAGCAAAAAACTGTAAGACATAAAGAATAGAGTGGAGCCGACTGAGAGATTA 1079

Qy   4756 AAATAAACTAGAAATATTTTATTAACACAGGCAATTTGAAATAATTTGTGCACTTCAGAATA 4815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1080 AAATAAACTAGAAATATTTTTATTAACACAGGCAATTTGAAATAATTTGTGCACTTCAGAATA 1139

Qy   4816 TTCTACAATAATATTATTTCCAATTTAATATCTTAAGAAATATCTTAAGAAATTATATG  4875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 TTCTACAATAATATTATTTCCAATTTAATATCTTAAGAAATATCTTAAGAAATTATATG  1199

Qy   4876 TAAGTACATGTGCATGTGTTTGAGGTAGGATATTTGAGGTTATTTTCTTTT        4935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1200 TAAGTACATGTGCATGTGTTTGAGGTAGGATATTTGAGGTTATTTTCTTTT         1259

Qy   4936 ATTCGGGTCAGGCAAAGCTTCTAAGGGATGTGAAAGGATATCCTCTTTCTCTTAGCTGA  4995
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 ATTCGGGTCAGGCAAAGCTTCTAAGGGATGTGAAAGGATATCCTCTTTCTCTTAGCTGA  1319

Qy   4996 GAGGAAGAGTGAGTTCTAAGTTAAAATATAATCAAGGAATTTCCCTGTCTTTGCTATTTGA 5055
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1320 GAGGAAGAGTGAGTTCTAAGTTAAAATATAATCAAGGAATTTCCCTGTCTTTGCTATTTGA 1379

Qy   5056 GATTGTGACCACAACAGGCGTTGCGTCGAAAGGGAAACTGAAAGGGCCGGGAGGGGAGGGAA 5115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1380 GATTGTGACCACAACAGGCCGTTGCCGTTGCTGGCTGAAAGGGAAACTGAAAGGGCGGGGAGGGGAGGGAA 1439

Qy   5116 ATAGATGAAAAACAAAAACAAAAACAAAAACTTCCCTAAGCAGCTCACAAAACATTTTAGC 5175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1440 ATAGATGAAAAACAAAAACAAAAACAAAAACTTCCCTAAGCAGCTCTACAAAACATTTTAGC 1499

Qy   5176 CCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGATACAAGGAAGTGTTA 5235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1500 CCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGATACAAGGAAGTGTTA 1559

Qy   5236 TGTAGCTGGAGCAGGGTGGACACTCATCAGTTCAGTTCAGTTACAAAAGTCCAGGCTGCT 5295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1560 TGTAGCTGGAGCAGGGTGGACACTCATCAGTTCAGTTCAGTTACAAAAGTCCAGGCTGCT 1619

Qy   5296 GAAATTAAACTCTGAGCCCATTCATGCCAGATCCAATCACGACAGAGATCAGAAGTTCA  5355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1620 GAAATTAAACTCTGAGCCCATTCATGCCAGATCCAATCACGACAGAGATCAGAAGTTCA  1679

Qy   5356 GAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTATACGTCAAGGACTCT 5415
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1680 GAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTATACGTCAAGGACTCT 1739

Qy   5416 GAAGCCGTGGAGAGGGGGAAGAACAACAACAGTAGAGGAGGATGCCCAGCTGGTAAGAATCGA 5475
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1740 GAAGCCGTGGAGAGGGGGAAGAACAACAACAGTAGAGGAGGATGCCCAGCTGGTAAGAATCGA 1799

Qy   5476 GTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAATCAAGAAAATCAGCGCTCCG 5535
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1800 GTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAATCAAGAAAATCAGCGCTCCG 1859

Qy   5536 CCTCTTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTATGTCTGATAGCATTT 5595
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1860 CCTCTTTGCAAATATGTATGAAGGAGAGAAGTGCCTAAACTTCTATGTCTGATAGCATTT 1919

Qy   5596 GACCCTATTGCTTTTAGCCTCCCGGCTTTATATCATATACACAGGTATTGTTGTATA  5655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1920 GACCCTATTGCTTTTAGCCTCCCGGCTTTATATCATATATACACAGGTATATGTGTATA 1979

Qy   5656 TTTTATATAATTGTTCTCCGT 5676
          |||||||||||||||||||||
Db   1980 TTTTATATAATTGTTCTCCGT 2000
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15779

Query Match             3.4%;  Score 190.6; DB 4;  Length 265038;
Best Local Similarity   74.8%;  Pred. No. 2.9e-33;
Matches  252;  Conservative    0;  Mismatches   84;  Indels    1;  Gaps    1;

Qy      3257 TAAACCTTTTAAATTTTTAAAAAATAGATGAGTGGTGGCTCATGCCTGTAATCCCA 3316
             |||||||||||| ||| ||| |||| || ||| |||||||||||||||||||||||
Db     31960 TAATTTATTTTATAAGATCAAAACCATGGCCAGGCGTGATGCTCACGCCTGTCATCCCCA 32019

Qy      3317 ACACTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCAGTCAGG 3376
             ||||||||| ||| |||||| |||||||||||||||||| |||||||||||||||
Db     32020 GCACTTTGGGAGGCCAAGGCAGGAGGAGATTGCTTGAGCCCAGGAGTTCGAGACCAGCCTGG 32079

Qy      3377 GCAACACAGCAAGACCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTATGG 3436
             ||||||| || |||||||| ||||| ||| || ||| ||||||| |||||||
Db     32080 GCAACATAGTGAGACCCTGTCTCTATAAAAAAAAATCAAGCAAAATTAGCCGGGTGTGA 32139

Qy      3437 TTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCC 3496
             |||||||| ||||||| ||| |||||||||||||||| || ||| |||||||
Db     32140 TGGCGCATGCCTGTGTAGTCCCAGCTACTCAAGAGGCTAAGGCGGGAGGATTGCTTGAGCCT 32199

Qy      3497 AGGAGGTTGAGGCTGCAGTGATCCATGACGCGCTGCTACACT-CAGTCTGGGTGACAGT 3555
             |||||||||||||||| |||||| |||||||||||||||||  ||||||||||||
Db     32200 GGGAAATTGAGGCTGCAGTGAGCCATGATCGTGCCACTGACTCCAGCCTGGGTGACAGA 32259

Qy      3556 GCAAGAAGCTGTCTCAAAAATAATAAATAAAAA 3592
             ||||| |||||||||||||||||||||| |||||
Db     32260 GCAAGACCCTGTCTCAAAAAAAAAGATAAAAA 32296
```

```
RESULT 3
US-09-659-791A-10
;  Sequence 10, Application US/09659791A
;  Patent No. 6383808
;  GENERAL INFORMATION:
;  APPLICANT: Brett P. Monia
;  APPLICANT: Susan M. Freier
;  TITLE OF INVENTION: ANTISENSE MODULATION OF CLUSTERIN EXPRESSION
;  FILE REFERENCE: RTS-0156
;  CURRENT APPLICATION NUMBER: US/09/659,791A
;  CURRENT FILING DATE: 2000-09-11
;  NUMBER OF SEQ ID NOS: 90
;  SEQ ID NO 10
;  LENGTH: 8133
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-659-791A-10

Query Match             3.3%;  Score 187.6; DB 3;  Length 8133;
Best Local Similarity   74.6%;  Pred. No. 3.3e-33;
Matches  249;  Conservative    0;  Mismatches   84;  Indels    1;  Gaps    1;

Qy      3264 TTTAAATTTTTTAAAAAATAGATGAGTGTGGCTCATGCCTGTAATCCCAACACTTT 3323
             |||||||    ||| || ||||| |||||||||||||||||||||||||||||||
Db      1057 TTTAAAGAGGAAAAAACTGAGCTGGGCATGGTGGCTCATGCCTGTGATGCCAGCACTTT 1116

Qy      3324 GGGAGGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCCAACAC 3383
             ||||||||  ||| |||| |||||||||| ||||||
Db      1117 GAGAGGCCAAGGCGGGAGGATCATTTGAGGCCAGGAGTTTGAGACCAGCCTGGGCAACAT 1176

Qy      3384 AGCAAGACCCCATATCTAAAAAAACAAAACAAAAATTACCTGGGTATGATGGTTGTGCT 3443
             || |||||||| ||||| ||| ||| |||||||||||| |||||
Db      1177 AGTGAGACCCTGTCTCTACAAAAATAAAACTTAAAAAATTAGCCGGGTGGTGGTGTGCA 1236

Qy      3444 CACCTGTAGTCCAAGCTCAGTGAGGCAGAAGGATCACTTGAGCCTGGGAGGT 3503
             ||||||||||||||||| ||| ||||||||| ||| || |||||||||||
Db      1237 CACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAGTCACCTGAGCCTGGGAAGTT 1296

Qy      3504 TGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAA 3562
             |||||||||||||| ||| |||| || |||||||  ||||||||||| |||||
Db      1297 GGAGGCTGCAGTGAGTGAGTATGATTGCACCATTGCATTCCAGCCTGGGCAACAAGAGTGAGAC 1356
```

```
Qy      3563 GCTGTCTCAAAAATAATAATAAAATAAAAAATAAC 3596
             ||||||||||||||| |||||| || |||||||||
Db      1357 CCTGTCTCTAAATTAAAATAAAAAATAATAAAATAAC 1390
```

```
RESULT 4
US-09-949-016-142852/c
;  Sequence 142852, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 142852
;  LENGTH: 601
;  TYPE: DNA
;  ORGANISM: Human
US-09-949-016-142852

Query Match             3.3%;  Score 186.4; DB 4;  Length 601;
Best Local Similarity   78.0%;  Pred. No. 2.1e-33;
Matches  237;  Conservative    0;  Mismatches   66;  Indels    1;  Gaps    1;

Qy      3290 GTGTGGTGGCTCATGCCTGTAATCCCAACACTTGGGAAGCCGGGTCGGGAGGATAGCTT 3349
             | |||||||||||||||||| ||||| |||||||||||| ||| |||||| ||||||||
Db       596 GCGGTGATGCTCACGCCTGTCATCCCAGCACTTTGGGAGGCCAAGGCAGGAGGATTGCTT 537

Qy      3350 GAGTCCAGCAGTTTGAGACCAGTCAGGCAACACAGCAACAGCCCCATATCTAAAAAAACA 3409
             |||  |||||||| |||||||||||| ||||| || |||||||| ||||| |||||| ||
Db       536 GAGCCCAGGAGTTCGAGACCAGCCTGGGCAACATAGTGAGACCCTGTCTCTATAAAAAAAA 477

Qy      3410 AAACAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAA 3469
             ||| ||| |||||||| |||||| ||| || ||| ||||||||| ||| ||||||||
Db       476 AAATCAAGCAAAATTAGCCGGGTGTGATGGCGCATGCCTGTGTAGTCCCAGCTACTCAAGAG 417

Qy      3470 GCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGACGCG 3529
             |||| || |||||||||||||||||| |||||||||||||||||||| |||||||
Db       416 GCTAAGGCGGGAGGATTGCTTGAGCCTGGGAAATTGAGGCTGCAGTGAGCCATGATCGTG 357

Qy      3530 CTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAATAAATA 3588
             ||||||||   ||||||||||| ||| ||||| |||||||||||||||| |||||| ||
Db       356 CCACTGACTCCAGCCTGGGTGACAGAGCAAGACCCTGTCTCTCAAAAAAAAAAAAAGATA 297

Qy      3589 AAAA 3592
             ||||
Db       296 AAAA 293
```

```
RESULT 5
US-09-949-016-17403
;  Sequence 17403, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
```

```
 ;   PRIOR APPLICATION NUMBER: 60/237,768
 ;   PRIOR FILING DATE: 2000-10-03
 ;   PRIOR APPLICATION NUMBER: 60/231,498
 ;   PRIOR FILING DATE: 2000-09-08
 ;   NUMBER OF SEQ ID NOS: 207012
 ;   SOFTWARE: FastSEQ for Windows Version 4.0
 ;   SEQ ID NO 17403
 ;   LENGTH: 17348
 ;   TYPE: DNA
 ;   ORGANISM: Human
US-09-949-016-17403

Query Match      3.3%;   Score 186;  DB 4;  Length 17348;
Best Local Similarity   74.3%;   Pred. No. 1.1e-32;
Matches 248;   Conservative   0;  Mismatches   85;  Indels   1;  Gaps   1;

Qy     3264  TTTAAATTTTTTAAAAAATAGATGAGTGGTGCTCATGCCTGTAATCCCAACACTTT 3323
             |||||     ||||||||||||||||||| ||| |||||||||||||||
Db     8546  TTTAAAGAGGAAAAAAACTGAGCTGGCTCACGTGCCTGTCATGCCTCAGCACTTT 8605

Qy     3324  GGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGGCCAACAC 3383
             |||||||||||||||||||
Db     8606  GAGAGGCCAAGGCCGGGAGGATCATTGAGGCAAGGATGGGGAGAA 8665

Qy     3384  AGCAAGACCCCATATCTCCTACACAGCAAGCTGAGGCAAGAGGATCACAGCAAGCTG 3443
             ||||||||
Db     8666  AGTGAGACCCTGTCTCCTACAAAAATAAAAACTTAAAAAATTAGCCGGGTGTGGTGCA 8725

Qy     3444  CACCTGTAGTCCAAGCTACACAGGATCAGTTGAGTCCAGCAGTTTGAGACCAGGCCAGGAGGT 3503
             |||||||||||||||||||||||||||||||||
Db     8726  CACCTGTAGTCTCAGCTACTCGGGAGGCTGAGGCAAGAGACCTGAGCCTGGAAGTT 8785

Qy     3504  TGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAA 3562
             ||||||||||||||||||||| ||||||||||||||
Db     8786  GGAGGCTGCAGTGAGCTATGATTGCACCATTGCATTCCAGCCTGGGCAACAGAGTGAGAC 8845

Qy     3563  GCTGTCTCAAAAATAATAAATTAAAAAATAATAAATAAAC 3596
             ||||||||||||||||||||||||
Db     8846  CCTGTCTGTAAATTAAAAAAATAATAAATAAAAATAAC 8879


RESULT 6
US-09-949-016-201827/c
 ;   Sequence 201827, Application US/09949016
 ;   Patent No. 6812339
 ;   GENERAL INFORMATION:
 ;   APPLICANT: VENTER, J. Craig et al.
 ;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ;   FILE REFERENCE: CL001307
 ;   CURRENT APPLICATION NUMBER: US/09/949,016
 ;   CURRENT FILING DATE: 2000-04-14
 ;   PRIOR APPLICATION NUMBER: 60/241,755
 ;   PRIOR FILING DATE: 2000-10-20
 ;   PRIOR APPLICATION NUMBER: 60/237,768
 ;   PRIOR FILING DATE: 2000-10-03
 ;   PRIOR APPLICATION NUMBER: 60/231,498
 ;   PRIOR FILING DATE: 2000-09-08
 ;   NUMBER OF SEQ ID NOS: 207012
 ;   SOFTWARE: FastSEQ for Windows Version 4.0
 ;   SEQ ID NO 201827
 ;   LENGTH: 601
 ;   TYPE: DNA
 ;   ORGANISM: Human
US-09-949-016-201827

Query Match      3.3%;   Score 185.6;  DB 4;  Length 601;
Best Local Similarity   74.0%;   Pred. No. 3.2e-33;
Matches 247;   Conservative   1;  Mismatches   85;  Indels   1;  Gaps   1;

Qy     3264  TTTAAATTTTTTAAAAAATAGATGAGTGGTGCTCATGCCTGTAATCCCAACACTTT 3323
             |||||     ||||||||||||||||||||||| |||||||||||||||||||
Db     466   TTTAAAGAGGAAAAAAACTGAGCTGGCTCACGTGCCTGTCATGCCTCAGCACTTT 407
```

```
Qy     3324  GGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGCCAACAC 3383
             |||||||||||||
Db     406   GAGAGGCCAAGGCGGGAGGATCATTGAGGCAAGGATGGGGAGAT 347

Qy     3384  AGCAAGACCCCATATCTCCTACACAGCAAGCTGAGGCAAGAGGATCACAGCAAGCTG 3443
             ||:|||
Db     346   AGTGAGACCCTGTCTCCTACAAAAATAAAAACTTAAAAAATTAGCCGGGTGTGGTGCA 287

Qy     3444  CACCTGTAGTCAAGCTACACAGGATCAGTTGAGTCCAGCAGTTTGAGACCAGGCCAGGAGGT 3503
             ||||||||||||||||||||||||||||||
Db     286   CACCTGTAGTCTCAGCTACTCGGGAGGCTGAGGCAAGAGGACCTGAGCCTGGAAGTT 227

Qy     3504  TGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAA 3562
             ||||||||||||||||| |||||||||||||||
Db     226   GGAGGCTGCAGTGAGCTATGATTGCACCATTGCATTCCAGCCTGGGCAACAGAGTGAGAC 167

Qy     3563  GCTGTCTCAAAAATAATAAATTAAAAAATAATAAATAAAC 3596
             |||||||||||||||||||||||
Db     166   CCTGTCTGTAAATTAAAAAAATAATAAATAAAAATAAC 133


RESULT 7
US-09-949-016-15382/c
 ;   Sequence 15382, Application US/09949016
 ;   Patent No. 6812339
 ;   GENERAL INFORMATION:
 ;   APPLICANT: VENTER, J. Craig et al.
 ;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ;   TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ;   FILE REFERENCE: CL001307
 ;   CURRENT APPLICATION NUMBER: US/09/949,016
 ;   CURRENT FILING DATE: 2000-04-14
 ;   PRIOR APPLICATION NUMBER: 60/241,755
 ;   PRIOR FILING DATE: 2000-10-20
 ;   PRIOR APPLICATION NUMBER: 60/237,768
 ;   PRIOR FILING DATE: 2000-10-03
 ;   PRIOR APPLICATION NUMBER: 60/231,498
 ;   PRIOR FILING DATE: 2000-09-08
 ;   NUMBER OF SEQ ID NOS: 207012
 ;   SOFTWARE: FastSEQ for Windows Version 4.0
 ;   SEQ ID NO 15382
 ;   LENGTH: 37292
 ;   TYPE: DNA
 ;   ORGANISM: Human
 ;   FEATURE:
 ;   NAME/KEY: misc_feature
 ;   LOCATION: (1)...(37292)
 ;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15382

Query Match      3.3%;   Score 185.4;  DB 4;  Length 37292;
Best Local Similarity   74.9%;   Pred. No. 2e-32;
Matches 245;   Conservative   0;  Mismatches   81;  Indels   1;  Gaps   1;

Qy     3284  AGATGAGTGGTGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGA 3343
             || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     15021 AGTTGAGTGTGTGGCTCACGTGCCTGTCACGCCTGAAATCCCAGCACTTTGGGAGGCCAAGGTGGGGAGAA 14962

Qy     3344  TAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGCCAACAGACCCCATATCTAAA 3403
             ||||||||||  ||||||||||||||||||||||||||
Db     14961 TTGCCTGAGGCCAGGAGTTCAAGACCAGCCTGGGCAACATAGCAAGACCCTATCTCTACA 14902

Qy     3404  AAAACAAAACAAAACAAAATTACCTGGTATGGTTGCTCACCTGTAGTCCAAGCTACA 3463
             ||||||||||||||||||||||||| ||
Db     14901 AAAATTAAAAAATAAAAAAGTAGCTGAGTGTGGTGACGTGCACCTGTGTCCCAACCACT 14842

Qy     3464  CAGGAAGCTGAGGCAGAAGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATG 3523
             || |||||| || |||||||||||||||||||||||||||||
Db     14841 CAGGAGGCTGAGGCAGGAGGATTGCTTGAGCCCAGGAGTTTGAGGCTGCAGTGAGCTATG 14782

Qy     3524  AACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGAGCTGTCTCAAAAATAATAAA 3582
             | | || |||||||||     ||  |||||| || |||
Db
```

```
Db  14781 ATCACAGCACTGCGCTCCAGCTTGGGTGACAGAGTGAGACCTTGTCTCTAAAAAAAAAG 14722

Qy   3583 TAAATAAAATAACTTTAAAAAACA 3609
          ||| |||||||   ||| ||
Db  14721 AAAAGAAAAAAAAAAAAAAGAAATGATRA 14695
```

RESULT 8
US-09-949-016-13444
; Sequence 13444, Application US/0994916
; Patent No. 6812339
; GENERAL INFORMATION:
..; APPLICANT: VENTER, J. Craig et al.
..; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
..; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
..; FILE REFERENCE: CL001307
..; CURRENT APPLICATION NUMBER: US/09/949,016
..; CURRENT FILING DATE: 2000-04-14
..; PRIOR APPLICATION NUMBER: 60/241,755
..; PRIOR FILING DATE: 2000-10-20
..; PRIOR APPLICATION NUMBER: 60/237,768
..; PRIOR FILING DATE: 2000-10-03
..; PRIOR APPLICATION NUMBER: 60/231,498
..; PRIOR FILING DATE: 2000-09-08
..; NUMBER OF SEQ ID NOS: 207012
..; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13444
; LENGTH: 172677
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13444

```
Query Match             3.3%;     Score 184.8;  DB 4;    Length 172677;
Best_Local_Similarity  69.6%;     Pred. No. 5.3e-32;
Matches  281;    Conservative   0;  Mismatches 117;  Indels  6;  Gaps  2;

Qy   3195 AAATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATTTAAATACGTTTAGGC 3254
Db  92789 AATTTAAAATATTACAAATCATACACAATAGATGTTGAGTTAAATTCAATTCAGTTGAC 92848

Qy   3255 TATAAACCTTTTAAATTTTTTAAAAAATGAT-----GAGTGTGGTGGCTCATGCCTGT 3309
Db  92849 TTTTTTGATTTGTATTTTTTTAAAAAAAAGTTGACCATGTGTGGTGGCTCGTGTCTGT 92908

Qy   3310 AATCCCAACACTTGGGAAGCCGGGTGCGGGGATAGCTTGAGGAGGAGTTCCAGCAGTTGAGACC 3369
Db  92909 AATCCCAGCACTTTGAAAGCTGAGTGGGAGGATAGCTTGAGGATGGATTAGCTTAAAAAATAAAAAATAAAAATAAGCTA 92968

Qy   3370 AGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAAACAAAACAAAATTACCTG 3429
Db  92969 AGCCTGGGCAACATGGCAACATGGCAAGACCTCATCTCTACTAAAAATAAAAAATAAGCTA 93028

Qy   3430 GGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAAGGATCACT 3489
Db  93029 GGCATAGCAGTGTGTTCATATAGTCCCAGCTACTCAGGAGGCACAGTGGGAGGATCGCT 93088

Qy   3490 TGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGATCATGAACGCGGCTGCTACACT-CAGTCTCGGG 3548
Db  93089 TGAACCCAGGAGGTTGAGGCTGCAGTGAGCTGTGATCTGAGTGGATCTGAGTGAACTGCACCTCCAGCCTAGG 93148

Qy   3549 TGACAGTGCAAGAAGCTGTCTCAAAATAATAAATAAAAA 3592
Db  93149 TGACAGAGAAAGACCTTGTCTCTCAAGAAATAAGCAAACAAAATAA 93192
```

RESULT 9
US-09-949-016-16949
; Sequence 16949, Application US/0994916
; Patent No. 6812339
; GENERAL INFORMATION:
..; APPLICANT: VENTER, J. Craig et al.
..; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
..; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
..; FILE REFERENCE: CL001307
..; CURRENT APPLICATION NUMBER: US/09/949,016
..; CURRENT FILING DATE: 2000-04-14
..; PRIOR APPLICATION NUMBER: 60/241,755
..; PRIOR FILING DATE: 2000-10-20
..; PRIOR APPLICATION NUMBER: 60/237,768
..; PRIOR FILING DATE: 2000-10-03
..; PRIOR APPLICATION NUMBER: 60/231,498
..; PRIOR FILING DATE: 2000-09-08
..; NUMBER OF SEQ ID NOS: 207012
..; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16949
; LENGTH: 33272
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16949

```
Query Match             3.3%;     Score 184.6;  DB 4;    Length 33272;
Best_Local_Similarity  68.6%;     Pred. No. 2.9e-32;
Matches 300;    Conservative   0;  Mismatches 129;  Indels  8;  Gaps  3;

Qy   3170 TTAGCTTAGTAACTTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAA 3229
Db  31344 TTGCTTTATAAGTTTTTCTCAATATTGTGAAAAAACTTCCACATCAGTAGATACATGTCC 31403

Qy   3230 ATGATATTATTTAAATACGTTAGGCTATAAAACCTTTTAAATTTTTTAAAAAAATAGATGA 3289
Db  31404 ATAATTTTTTTTTGTATTGTTTGATTTTA---TTAAATTTTATTGAAATATAGGCTGG 31460

Qy   3290 GTGTGGTGGCTCATGCCTGTAATCCCAACACTTGGGAAGCCGGGTGCGGGTGGAGATAGCTT 3349
Db  31461 GTGCAGTGGCTTACCCTCTGTAATCCCAACACTTTGAGAGGCTGAGGTGGGATGGATTCCTT 31520

Qy   3350 GAGTCCAGCAGTTTGAGACAGTTTAAATAACAGCGAGCCCCATATCTAAAAAAACA 3409
Db  31521 GAGCTCAGGAGTTCAAGACCACCACCCTGGCAACACAGCGAGTACTATCTCTACAAAAAAA 31580

Qy   3410 AAACAAAACAAAATTACCTGGGATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAA 3469
Db  31581 ATAGA---AAAATTAGCTGGGCATGGTGGCGTGTGCCTGTGGTCCCAGCTACTCGGGAG 31636

Qy   3470 GCTGAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGAGCCG 3529
Db  31637 GCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATTGCA 31696

Qy   3530 CTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAATAAATA 3588
Db  31697 CTGCTGCACTCCAGCGTGGGTGACAGAGCAAGACCCCTGTCTCAAAAACCCCAAAAACAA 31756

Qy   3589 AAAATAACTTTTAAAAA 3605
Db  31757 AACTGAAATATCAAATA 31773
```

RESULT 10
US-09-949-016-58699
; Sequence 58699, Application US/0994916
; Patent No. 6812339
; GENERAL INFORMATION:
..; APPLICANT: VENTER, J. Craig et al.
..; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
..; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
..; FILE REFERENCE: CL001307
..; CURRENT APPLICATION NUMBER: US/09/949,016
..; CURRENT FILING DATE: 2000-04-14
..; PRIOR APPLICATION NUMBER: 60/241,755
..; PRIOR FILING DATE: 2000-10-20
..; PRIOR APPLICATION NUMBER: 60/237,768
..; PRIOR FILING DATE: 2000-10-03
..; PRIOR APPLICATION NUMBER: 60/231,498
..; PRIOR FILING DATE: 2000-09-08
..; NUMBER OF SEQ ID NOS: 207012
..; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 58699
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58699

Query Match   3.2%;   Score 184.4; DB 4;   Length 601;
Best Local Similarity 69.3%;   Pred. No. 6.1e-33;
Matches 280;   Conservative 1; Mismatches 117; Indels 6; Gaps 2;

Qy   3195 AAATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATTTAAATACGTTAGGC 3254
Db   156 AATTTAAAAATTATTACAAATCATACATACAATAGATGTTGAGTTAAATTCAATTCAGTTGAC 215

Qy   3255 TATAAACCTTTTAAATTTTTTAAAAAAATAGAT-----GAGTGTGGTGGCTCATCGCCTGT 3309
Db   216 TTTTTTGATTTGGTATTTTTTTAAAAAAAAGTTGACCATGTGTGGTGGCTCGTGTCTGT 275

Qy   3310 AATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACC 3369
Db   276 AATCCCCAGCACTTTGGGAAAGCTGAGYTGGAAGGATAGCTTGAGACCAGGAGTTTGAGACC 335

Qy   3370 AGTCAGGGCAACACAGCAAGACCCCATATCTAAAAAAACAAAACAAAAATTACCTG 3429
Db   336 AGCCTGGCAACATGCAAGACCTCATCTCTACTAAAAATAAAAATAAAAATAAGCTA 395

Qy   3430 GGTATGGTTGGTGTGCTCACCTGGTCCAAGCTACACAGGAGCTGAGGCAGCAGAAGGATCACT 3489
Db   396 GGCATAGCAGTGTGTTCATATAGTCCCAGCTACTCAGGAGGCAGCAGGCCAGGTGGGAGGATCGCT 455

Qy   3490 TGAGCCCAGGAGGTTGAGGCTGCAGTGAGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGG 3548
Db   456 TGAACCCAGGAGGTTGAGGGCTGCAGTGAGCTGCAGTGATCCACTCCAGCCTAGG 515

Qy   3549 TGACAGTGCAAGAAGCTGTCTCCAAAAATAATAAAATAAATAAAAA 3592
Db   516 TGACAGAGAAAGACCCTTGTCTCAAGGAAATAAGCAAACAAAATAA 559

RESULT 11
US-09-949-016-132156
; Sequence 132156, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 132156
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-132156

Query Match   3.2%;   Score 184; DB 4;   Length 601;
Best Local Similarity 62.4%;   Pred. No. 7.5e-33;
Matches 322;   Conservative 1; Mismatches 186; Indels 7; Gaps 2;

Qy   3195 AAATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATTTAAATACGTTAGGC 3254
Db   5 ACACAACAACAACAAAAAGATAAACTGGACTTTATTAAGTTAAAAACTTGAACTTCAA 64

Qy   3255 TATAAACCTTTTTAAATTTTTTAAAAAAATAGATAGAGTGTGGTGGCTCATCGCCTGTAATCC 3314

Db   65 AAGATACCATAAGAAAAATGAAAAAGGCAGGCAGGCCAGTGCAGTGGCTCACGCCTGTAATCC 124

Qy   3315 CAACACTTTGGGAGGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGACGACCAGTCA 3374
Db   125 CACCACTTTGGGAGGCCGAGGCGGGGGATTGCTTGAGGCCAGGAGTTCGAGACCAGTCT 184

Qy   3375 GGGCAACAGCAGCAAGACCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTAT 3434
Db   185 GGCCAACATGGCAAAACCCCTGTCTCT-----ACTAAAAATACAAAAATTAGCTGGGTGT 238

Qy   3435 GGTTGTGCTCACCTGGTAGTCACAGGAGGCTGAGGCAGGAGGATCACTTGAGC 3494
Db   239 GGTGGCGGGCACCTGTAATCCCAGTTACTCCAGGAGGCTGAGGCAGGAGAATCGCTTGAAC 298

Qy   3495 CCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACA 3553
Db   299 CCRGGAGGCCAGAGGTTGCAGTGAGCCAAGATCACCGCCGTTGCACTCCAGCGGCCGACA 358

Qy   3554 GTGCAAGAAGCTGTCTCAAAAATAATAAATAAATAAAAAATAACTTTTAAAAACAAAAT 3613
Db   359 GAGCAAAACTCCATCTCAAAAAAAAAAAAAAAAGAAAATGAAAAAGAGACAAAC 418

Qy   3614 TAATTAAATTTTAAAACACACACCACTAGAGATGTTTGCAAATGTTGATTGATTTGGGAGTC 3673
Db   419 TCAATGAATAGGATAAAATATTTGCAAATCATATATCTGATAAGGGACCTGTATCTAGAA 478

Qy   3674 TATATCCCTGGAAGTTAATTTAAAATATTTAGAAGA 3709
Db   479 TATAAAAAGAATTCTTGTAACTTGTAACTCAATAATAAGAAGA 514

RESULT 12
US-09-949-016-119415
; Sequence 119415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119415

Query Match   3.2%;   Score 183.4; DB 4;   Length 601;
Best Local Similarity 62.2%;   Pred. No. 1e-32;
Matches 304;   Conservative 1; Mismatches 182; Indels 2; Gaps 1;

Qy   3143 TTTGGTTGCCATTTAATTTCTGTTCCTCCTTAGCTTAGCTTAGTAACTTAGGATTTTAAATAACA 3202
Db   47 TTTTTTTCTTTTTTTTGCTTTATTTGTCTTCTAAATTTCTATATTCTAAGTGTTCTATAATC 106

Qy   3203 ACTATTGAAATCATGACATACGTTTAAATGATATTAGTATTAAAATACGTTAGGTATAAACC 3262
Db   107 CTTTTATAATTAAAGAAAAATTTTAAGTTTTATTTCAGGACACCATTGAATTAATAC 166

Qy   3263 TTTTAAATTTTTTAAAAAAATAGATAGATAGAGTGTGGTGGCTCATCGCCTGTAATCCCAACACTT 3322
Db   167 TGAATAAATTAACTAAATAGATAGTGGTATAGTGCTATATCCCAGCACTT 226

```
Qy  3323  TGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACA  3382
          ||||| ||| ||  |  ||||||||  ||||| |||||| ||||||||| |  ||||||||
Db   227  TGGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGGTGACA  286

Qy  3383  CAGCAAGAGACCCCCATATCTAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGC  3442
          ||| |||| |  ||  | ||||||||  ||| |||| ||||   |
Db   287  TAGTGAGATCCTTGTMTCTGCAAAATAAAAAAAAAATGTTTAAATCAGCCAGGTGTGGTGGTGC  346

Qy  3443  TCACCTGTAGTCCAAGCTACACAGGAGCTGAGGCGAAGGATCACTTGAGCCCAGGAGG  3502
          |||||||||||   |||||| |  ||| |||||| ||||| ||||
Db   347  CTGCCCTGTAGTTTCAGATATGTTGAAGGATTGCTTGAGTGGTTGCCGGGAGT  406

Qy  3503  TTGAGGCTGCAGTGATCCATTGAACGCGCTGCTACACT--CAGTCTGGGTGACAGTGCAAG  3560
          |||||||||||| |||| | |||  || |  ||| |||  |||||||||||| ||||||||
Db   407  TTGAGGCTGCAGTATGATCGTCATGCCATGCCATGCCATTGCCATTGCCCTTTGGGTGACAGAGCAAG  466

Qy  3561  AAGCTGTCTCAAAAATAAAATAAATAATAAAATAAACTTTTAAAAAACAAAAAT TAATTAA  3620
          | |||||||||  | |||||  | | ||| |||||||||
Db   467  ACCCTGTCTCAACAAAAAACAGAGAAGTATGTATATAATGATATATAGTCCAGTATTATTTA  526

Qy  3621  ATTTTAAAA  3629
          |||||
Db   527  ACAGTAATA  535
```

RESULT 13
US-09-949-016-12256/c
; Sequence 12256, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12256
; LENGTH: 36228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12256

Query Match   3.2%; Score 182.8; DB 4; Length 36228;
Best Local Similarity   62.4%; Pred. No. 7.9e-32;
Matches 322; Conservative 0; Mismatches 187; Indels 7; Gaps 2;

```
Qy  3195  AAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATTAAATACGTTAGGC  3254
          | |||||| |||  |  |||||||||||||||||| |||| ||||| || ||||
Db  7176  ACACAACAACAACAACAAAAGATAACAAACTTGGACTTTAAAAAATAGATTTAATACTTGAACTTCAA  7117

Qy  3255  TATAAACCTTTTAAATTTTTAAAAAATGAGTGTGGCTCATGCCTGTAATCC  3314
          |||  |   |||| |      ||||| |  |||  ||  | ||| ||||||
Db  7116  AAGATACCATAAGAAAAATGAAAAAGGCAGGCGGGGATTGCTTGAGGCCAGGAGTTCGAGACCAGTCT  7057

Qy  3315  CAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTCA  3374
          |||||||||||||| |||||  ||||  || ||||||  ||||
Db  7056  CACCACTTTGGGAGGCCGGGCGGGGGGGATTGCTTGAGGCCAGGAGTTCGAGACCAGTCT  6997

Qy  3375  GGGCAACAGCAGCAAGACCCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTAT  3434
          ||||||| |||||||| || |||  ||  |||||| ||||  |
Db  6996  GCCCAACATGGCAAAACCCTGTCTCT-----ACTAAAAATACAAAAATTAGCGTGGTGT  6943

Qy  3435  GGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAAGGATCACTTGAGC  3494
          ||| | ||  ||||||| | |||| ||| |||||||||||||||| ||| ||  |||| |
Db  6942  GGTGGCGGGCACCTGTAATCCCAGTTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAAC  6883

Qy  3495  GTGCAAGAAGCTGTCTCAAAAATAATAAATAATAAATAACTTTTAAAAAACAAAAAT  3553
          ||  || |
Db  6882  CCGGGAGGCCAGGAGGTTGCAGTGAGCCAAGATCACGCCCGTTGCACTCCAGCCAGGGCGACA  6823
```

RESULT 14
US-09-949-016-15468/c
; Sequence 15468, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15468
; LENGTH: 36228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15468

Query Match   3.2%; Score 182.8; DB 4; Length 36228;
Best Local Similarity   62.4%; Pred. No. 7.9e-32;
Matches 322; Conservative 0; Mismatches 187; Indels 7; Gaps 2;

```
Qy  3195  AAATAACAACTATTGAAATCATGACATACGTTTAAATGATATATTAAATACGTTAGGC  3254
          | |||||| |||  |  |||||||||||||||||| |||| ||||| || ||||
Db  7176  ACACAACAACAACAACAAAAGATAACAAACTTGGACTTTAAAAAATAGATTTAATACTTGAACTTCAA  7117

Qy  3255  TATAAACCTTTTAAATTTTTAAAAAATGAGTGTGGCTCATGCCTGTAATCC  3314
          |||  |   |||| |      ||||| |  |||  ||  | ||| ||||||
Db  7116  AAGATACCATAAGAAAAATGAAAAAGGCAGGCGGGGATTGCTTGAGGCCAGGAGTTCGAGACCAGTCT  7057

Qy  3315  CAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTCA  3374
          |||||||||||||| |||||  ||||  || ||||||  ||||
Db  7056  CACCACTTTGGGAGGCCGGGCGGGGGGGATTGCTTGAGGCCAGGAGTTCGAGACCAGTCT  6997

Qy  3375  GGGCAACAGCAGCAAGACCCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTAT  3434
          ||||||| |||||||| || |||  ||  |||||| ||||  |
Db  6996  GCCCAACATGGCAAAACCCTGTCTCT-----ACTAAAAATACAAAAATTAGCGTGGTGT  6943

Qy  3435  GGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAAGGATCACTTGAGC  3494
          ||| | ||  ||||||| | |||| ||| |||||||||||||||| ||| ||  |||| |
Db  6942  GGTGGCGGGCACCTGTAATCCCAGTTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAAC  6883

Qy  3495  CCAGGAGGTTGAGGCTGCAGTGATCATGAACGCGCTGCTACACT-CAGTCTGGGTGACA  3553
          || |||||  | |||||||||||  ||  ||| |  ||| |||| |||| ||||||||||
Db  6882  CCGGGAGGCCAGGAGGTTGCAGTGAGCCAAGATCACGCCCGTTGCACTCCAGCCAGGGCGACA  6823

Qy  3554  GTGCAAGAAGCTGTCTCAAAAATAATAAATAATAAATAACTTTTAAAAAACAAAAAT  3613
          | ||| |  ||| ||||||||||  |  |
Db  6822  GAGCAAAACTCCATCTCAAAAATAAAAAAAATGAAAAATGAAAAGAAGACAAAC  6763

Qy  3614  TAATTAAATTTTAAAAACACACACACTAGAGATGTTTGCAAATTGATTATTTGGGAGTC  3673
          | |
Db  6762  TCAATGAATAGGATAAAATATTTGCAAATCATATATCTGATAAGGGACCTGTATCTAGAA  6703

Qy  3674  TATATCCCTGGAAGTTAATTTAAAATATTTAGAAGA  3709
          |||  |
Db  6702  TATAAAAAGAATTCTTGTAACTCAATAATAGAGAAGA  6667
```

```
Db      6822  GAGCAAAACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAGAAAAAATGAAAAGACAAC  6763

Qy      3614  TAATTAAATTTAAAAACACAACACACTAGAGATGTTGCAAATTGATTATTTGGGAGTC  3673
               |||| |       |||
Db      6762  TCAATGAATAGGATAAAATATTTGCAAATCATATATCTGATAAGGGACCTGTATCTAGAA  6703

Qy      3674  TATATCCCTGGAAGTTAATTTAAAAATATTTAGAAGA  3709
               ||||   ||  |   ||||| | ||||||
Db      6702  TATAAAAGAGAATTCTTGTAACTCAATAATAAGAAGA  6667
```

RESULT 15
US-09-949-016-12672/c
; Sequence 12672, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 12672
; LENGTH: 134140
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(134140)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12672

Query Match   3.2%;   Score 182.6; DB 4;   Length 134140;
Best Local Similarity 64.4%; Pred. No. 1.5e-31;
Matches 289; Conservative 0; Mismatches 159;   Indels 1;   Gaps 1;

```
Qy      3268  AATTTTTAAAAAAATAGATGAGTGTGGTGCTCATGCCTGTAATCCCAACACTTTGGGA  3327
               ||| |      ||||||||| |||||    ||| ||| |||||  |||
Db     94152  AAAATGTACAAGAAAATTATGGCCGGGTGTAGTGCCTGTAATTCCAGCACTTAGGGA  94093

Qy      3328  AGCCGGTGCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCAGTCAGGCAACACAGCA  3387
               |||||    ||| |||||  |||| |||||||  ||||    | ||||
Db     94092  AGCCAAGATGGGAGTATCACTTAAGCCCAGGAGTTCAGGATCAGCCTAGACAACATGGAG  94033

Qy      3388  AGACCCCATATCTAAAAAACAAAACAAAACAAAATCAAAAATTAACCTGGGTATGGTTGTGCTCACC  3447
               |||||          ||||||||||| |||||||| |||| ||||||
Db     94032  AAATCCTCTCCTCACAAAACAAAAATACAAAAAATTAGCTGGGTGTGATGGTACACGCC  93973

Qy      3448  TGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGCTCACTTGAGCCCAGGAGGTTGAG  3507
               |||||||||||||| |||| |||||||||||| |||||| |||||
Db     93972  TGTAGTCCCAGCTACTTAGGAGGCTGAGGTGGGAGGATCACTTGAGCCCAGGAAGTTGAG  93913

Qy      3508  GCTGCAGTGATCCATGAACGCGCTGCTACACTC-AGTCTCGGTGACAGTGCAAGAAGCTG  3566
               |||||||    |||||| |||||||||||| |  ||||  | |||
Db     93912  ACTGCAGTGATCCATGATTGTGCCACTGCACCTTGGTCACCTAGCCCTGGTGACAACATAAGACCCTG  93853

Qy      3567  TCTCAAAAATAATAATAAATAATAATAAAAAATAACTTTAAAAAACAAAAATTAATTAAATTTTA  3626
               |||||     |||    ||||| | | ||| ||||   |  ||||  ||| |  ||||
Db     93852  TTTCCAAAAGAAAAGAAAGAAATTACTTTTCAACTATGAAGTGGAATTAT  93793

Qy      3627  AAAACACAACACACTAGAGATGTTTGCAAATTGATTATTTGGGAGTCTATATCCCTGGAA  3686
               ||||||  |||| ||||||||||||||||||||| || || |||  |||  |  | |||
Db     93792  TAGACACCAACCAATTAACTGTATGAGATTGTAGGAGATAAAATTTATGTTGAGTTAAAA  93733

Qy      3687  GTTAATTTAAAATATTTAGAGAGAGTTCTT  3715
```

```
Db     93732  TAAAAATTAAATCAGAAGAAGAAAATTGTTT  93704

Search completed: June 23, 2005, 14:09:42
Job time : 861 secs
```

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    June 23, 2005, 00:34:38 ; Search time 3068 Seconds
                                      (without alignments)
                           11563.865 Million cell updates/sec

Title:           US-10-030-294-1
Perfect score:   5676
Sequence:        1 cagctgttcaggatgctg.........tttatataattgttcccgt 5676

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        6067389 seqs, 3125258755 residues

Total number of hits satisfying chosen parameters:      12134778

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications NA:*
         1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
         2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
         3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
         4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
         5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
         6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
         7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
         8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
         9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
        10: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
        11: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
        12: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
        13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
        14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
        15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
        16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
        17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
        18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
        19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
        20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
        21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
        22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
        23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
        24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
        25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
        26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
c 1	196.2	3.5	65793	21	US-10-703-817-3	Sequence 3, Appli
c 2	196	3.5	304905	18	US-10-271-416-1	Sequence 1, Appli
c 3	193	3.4	13070	21	US-10-741-600-17605	Sequence 17605, A
c 4	193	3.4	156318	21	US-10-741-600-17574	Sequence 17574, A
5	188.4	3.3	135005	20	US-10-723-860-2320	Sequence 2320, Ap
6	187.6	3.3	8133	18	US-10-380-124-10	Sequence 10, Appl
7	186	3.3	18861	11	US-09-984-429-513	Sequence 513, App
c 8	185.6	3.3	611	13	US-10-027-632-181221	Sequence 181221,
c 9	185.6	3.3	611	17	US-10-027-632-181221	Sequence 181221,
c 10	185.4	3.3	3287	13	US-10-027-632-115440	Sequence 115440,
c 11	185.4	3.3	3287	17	US-10-027-632-115440	Sequence 115440,
c 12	183	3.2	335199	21	US-10-496-011-4	Sequence 4, Appli
13	182.6	3.2	47903	19	US-10-322-281-808	Sequence 808, App
14	182.4	3.2	736	13	US-10-027-632-17813	Sequence 17813, A
15	182.4	3.2	736	17	US-10-027-632-17813	Sequence 17813, A
16	182.2	3.2	36651	9	US-09-964-469-3	Sequence 3, Appli
17	182.2	3.2	36651	16	US-10-425-962-3	Sequence 3, Appli
18	182	3.2	55735	21	US-10-741-600-17698	Sequence 17698, A
c 19	181.8	3.2	50000	17	US-10-364-505-8	Sequence 8, Appli
c 20	181.8	3.2	50000	19	US-10-681-199-8	Sequence 8, Appli
21	181.4	3.2	145806	20	US-10-719-993-6943	Sequence 6943, Ap
22	181	3.2	58985	10	US-09-901-152-3	Sequence 3, Appli
23	181	3.2	58985	21	US-10-483-593-3	Sequence 3, Appli
24	181	3.2	143601	10	US-09-855-824-3	Sequence 3, Appli
25	181	3.2	143601	20	US-10-476-542-3	Sequence 3, Appli
c 26	180.8	3.2	2214	13	US-10-027-632-110324	Sequence 110324,
c 27	180.8	3.2	2214	17	US-10-027-632-110324	Sequence 110324,
28	180.6	3.2	198522	13	US-10-087-192-244	Sequence 244, App
29	180.6	3.2	13487	20	US-10-719-993-6919	Sequence 6919, Ap
30	180.2	3.2	95832	21	US-10-741-600-17662	Sequence 17662, A
31	180	3.2	94720	18	US-10-052-482-160	Sequence 160, App
32	179.8	3.2	108316	17	US-10-292-798-1789	Sequence 1789, Ap
33	179.8	3.2	108317	15	US-10-017-161-2143	Sequence 2143, Ap
34	179.6	3.2	51657	17	US-10-057-475B-10475	Sequence 10475, A
35	179.6	3.2	51657	17	US-10-154-884B-10475	Sequence 10475, A
c 36	179.6	3.2	65608	9	US-09-962-436-292	Sequence 292, App
c 37	179.6	3.2	65608	9	US-09-962-832-119	Sequence 119, App
c 38	179.6	3.2	65608	9	US-09-954-531-180	Sequence 180, App
c 39	179.6	3.2	65608	21	US-10-843-641A-1247	Sequence 1247, Ap
c 40	179.6	3.2	65608	21	US-10-843-641A-2751	Sequence 2751, Ap
c 41	179.6	3.2	65608	21	US-10-843-641A-6005	Sequence 6005, Ap
c 42	179.4	3.2	68744	20	US-10-719-993-6752	Sequence 6752, Ap
c 43	179.4	3.2	160556	20	US-10-719-993-6827	Sequence 6827, Ap
c 44	179.2	3.2	812	13	US-10-027-632-170028	Sequence 170028,
c 45	179.2	3.2	812	17	US-10-027-632-170028	Sequence 170028,

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

```
RESULT 1
US-10-703-817-3/c
; Sequence 3, Application US/10703817
; Publication No. US2005011817A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: BRAUN, ANDREAS
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF MELANOMA AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4061-UT
; CURRENT APPLICATION NUMBER: US/10/703,817
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: 60/489,703
; PRIOR FILING DATE: 2003-07-23
; PRIOR APPLICATION NUMBER: 60/424,475
; PRIOR FILING DATE: 2002-11-06
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.3
;
; SEQ ID NO 3
; LENGTH: 65793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-703-817-3

  Query Match           3.5%;  Score 196.2;  DB 21;  Length 65793;
  Best Local Similarity   70.0%;  Pred. No. 4e-30;
  Matches  278;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;
```

Qy   3262 CTTTTAAATTTTTTAAAAAAATAGATGAGTGTGGTGGCTGCTCATGCCTGTAATCCCAACACT 3321
           |||||||||||| ||||||| ||||||||||| |||||||||||||||||||||||||||
Db   18059 CTATAAAGTTTTTTACAGGGAAGGAAAGCCAGGTGCGGTGGCTGCGTCACACCTGTAATCCCAGCACT 18000

Qy   3322 TTGGGAAGGCCGGGTCGGGAGGATAGCTTGAGTCAGCAGTTTGAGACCAGTCAGGGCAAC 3381
           ||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db   17999 TTGGGAAGCTGAGGCAGGTCAGTGAGCTGATTGTCTTGAGCTCAGGAGTTTGAGCCAGGGCAAC 17940

Qy   3382 ACAGCAAGACCCCATCTCTAAAAAAAACAAAACAAAATTACCTGGGTATGGTGGTGTG 3441
           ||||||||| ||||||| |||||||||||||||||||| |||||||||||||| |||
Db   17939 ATGGCAAAAAACCCATCCTCTACATAAAATACCAGAAAAAAAAAAAATTAGCTGCGTGGTGGTG 17880

Qy   3442 CTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAG 3501
           |||||| |||||||||||| |||| ||||||||||||| |||||||||||||||||||||
Db   17879 CACACCCGTGGTCCTGCTTACTCAGGAGGCTGAGGCAGGAGGATGGGAGATTGCTTGAGCCCAGGAG 17820

Qy   3502 GTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAG 3560
           |||||||||||||||||||||||| |||| ||||||| |||||||||||||||| ||||
Db   17819 ATTGAGGCTGCAGTGAGCCAAGATCATGCCCACTGTACTCCCAGCCTGGGCAACAGAGCAAG 17760

Qy   3561 AAGCTGTCTCAAAAATAATAATAAAAATAACTTTTAAAAAAACAAAAATTAATTAA 3620
           ||||||||| ||||||||||| ||| |||||||||||||||||||| ||||||||||
Db   17759 ACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAATAATATATATATATATATATATA 17700

Qy   3621 ATTTTAAAAACAACACACACTAGAGATGTTTGCAAAT 3657
           |||||||| |||||| |||||| |||||||||||
Db   17699 TATAATATATACACACACATATTATGTGTATATAT 17663

RESULT 2
US-10-271-416-1/c
; Sequence 1, Application US/10271416
; Publication No. US20040043021A1
; GENERAL INFORMATION:
;   APPLICANT: Little, Tim
;   APPLICANT: Little, Randall D.
;   APPLICANT: Van Eerdewegh, Paul
;   APPLICANT: Dupuis, Josee
;   APPLICANT: Del Mastro, Richard G.
;   APPLICANT: Allen, Kristina
;   TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES
;   TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY
;   FILE REFERENCE: 2976-4045
;   CURRENT APPLICATION NUMBER: US/10/271,416
;   CURRENT FILING DATE: 2002-10-11
;   PRIOR APPLICATION NUMBER: 60/328,424
;   PRIOR FILING DATE: 2001-10-11
;   NUMBER OF SEQ ID NOS: 9
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 304905
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-271-416-1

  Query Match   3.5%; Score 196; DB 18; Length 304905;
  Best Local Similarity 73.3%; Pred. No. 9.5e-30;
  Matches 264; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

Qy   3279 AAAATAGATGAGTGTGGTGGCTGCTCATGCCTGTAATCCCAACACTTTGGGAGGCCGGGTCGG 3338
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 156005 AATCTCGTGGGGTGCGGTGGCTAGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGG 155946

Qy   3339 GAGGATAGCTTGAGTCAGCAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCATGT 3398
           ||||||||| |||||||||||||||||| |||||||||||||||||||||| |||||||
Db 155945 GAGGATTGCTTGAGCCCAGGAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCATGT 155886

Qy   3399 CTAAAAAAACAAAACAAAACAAAATTACCTGGGTGTGTGGTGCGTGCGTGTGTAGTCCAAG 3458
           ||||||||||||||| |||||||||||||||||| |||| |||||||||||||||||||
Db 155885 CTACAAAAAAAAAAAATCAAAAACGTTAGCCAGGCGTGGTGTGCAAACCCTGTAGTTTCAG 155826

Qy   3459 CTACACAGGAAGCTGAGGCAGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGCTGAGTGAT 3518
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||

```
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17574
;   LENGTH: 156318
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(156318)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17574

Query Match             3.4%;  Score 193;  DB 21;  Length 156318;
Best Local Similarity  70.8%;  Pred. No. 2.9e-29;
Matches   271;  Conservative    0;  Mismatches  110;  Indels    2;  Gaps    1;

Qy       3210 AAATCATGACATACGTTAAATGATATTATTTAAATACGTTAGGCTATAAACCTTTTAAA 3269
              | ||||   ||||||  |||| | |||||| |||| |||||  |||||
Db      66635 ATATCTGTACAAAACATCTGATTTCACTTTTCAAAATCTTTAGGTAGTAAAGTTAAAATC 66576

Qy       3270 TTTTTAAAAAAATAGATGAGTGTGGTGCTCATGCCTGTAATCCCAACACTTTCGGGAAG 3329
              | ||||||        |||||||| | ||| |||||||||  ||||||||
Db      66575 TAAAGGTCAAAGATTGCTGCTGGTATGGTAGCTAGTCACGCCTGTAATCCCAACACTCTAGGAGG 66516

Qy       3330 CCGGGTCGGGAGGATAGCTTGAGTCCAGGAGTTTGAGACCAGTCAGGCCAACACAGCAAG 3389
              |||||  |  |  ||||||    |||||||||||||||||||||||  ||| |
Db      66515 CCGAGGCGGGCATATCGCTTGAGCCCAGGAGGTTTGAGACCAGCCAACAAGGCAA 66456

Qy       3390 ACCCCATATCTAAAAAACAAAAACAAAACAAAATTACCTGGGTATGGTTGCTCACCTG 3449
              |||||||||   ||| |||||| |  | ||||||  ||||||||||||||||
Db      66455 ACCCCATATCTTACAAAATAAAAAATACA--ATAATGAGCTGGGCATGGTGGCACGGCACCTG 66398

Qy       3450 TAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGTTGAGGC 3509
              |||||| |||  |||||||| |||  |||||||
Db      66397 TGGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGGATGGCTTGAGCCCAGGAGGTTGGGGC 66338

Qy       3510 TGCAGTGATCCATGAAACGCGCCGCTGCTACACTCAGTCTCGGGTGACAGTGCAAGAAGCTGTCT 3569
              ||||||||||    ||||||||||||| |||||||||||||||||||
Db      66337 TGCAGTGAGCTGTGACTGTACTACTGCACTCCACCTAGGTGACAGAATAGCACCCCTGTCT 66278

Qy       3570 CAAAAAATAATAAATAAATAAAA 3592
              |||||||||||| ||| ||||||
Db      66277 CAAAAAAACAAATAAAAATAA 66255


RESULT 5
US-10-723-860-2320
; Sequence 2320, Application US/10723860
; Publication No. US2004025360A1
; GENERAL INFORMATION:
;   APPLICANT: Aziz, Natasha
;   APPLICANT: Ginsburg, Wendy M.
;   APPLICANT: Zlotnik, Albert
;   TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;   TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
;   FILE REFERENCE: 05882.0193.NPUS01
;   CURRENT APPLICATION NUMBER: US/10/723,860
;   CURRENT FILING DATE: 2003-11-26
;   PRIOR APPLICATION NUMBER: 60/429,739
;   PRIOR FILING DATE: 2002-11-26
;   NUMBER OF SEQ ID NOS: 8393
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2320
;   LENGTH: 135005
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-2320

Query Match             3.3%;  Score 188.4;  DB 20;  Length 135005;
Best Local Similarity  71.8%;  Pred. No. 2.6e-28;
Matches   260;  Conservative    0;  Mismatches  101;  Indels    1;  Gaps    1;
```

```
Qy       3263 TTTTAAATTTTTTAAAAAAATAGATGAGTGTGGTGGCTCATGCCTGTAATCCCAACACTT 3322
              |||||
Db      26435 TTTTATAAAATGGTGACAGTGGCTGGCTGCAGTGCGGCTCATGCCTGTAATCCCAACACTT 26494

Qy       3323 TGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCAGTCAGGGCAACA 3382
              ||||| ||||||||| || |||  |||| ||||||||| |||||||
Db      26495 TGGGAGGCTGAGGTCAGGAGGGATTGCTTGAGCCCATGAGTTTGAGACCAGCCTGGCCAACA 26554

Qy       3383 CAGCAAGACCCCATATCTAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTGTGC 3442
              | | |||||||||   ||||||||||||||||||||
Db      26555 GGGTAAAATCCCATTCCATCTCAACACAAAAATACACAACAAAAAATTTAGCTGGGCATGGTCGTGT 26614

Qy       3443 TCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGG 3502
              || |||||||||||||||||||||||||||||||||||||
Db      26615 ACATCTGTGGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGGATCACCTGGGTGACAGTGCAAGA 26674

Qy       3503 TTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGA 3561
              | ||||  |||||||||   |||||||||||||
Db      26675 TCTAGACTGCGCGAGCGCTGTGATTGTGCCACTCGCACTCCCAGTCTCGGGCAACAGAGTGAGA 26734

Qy       3562 AGCTGTCTCAAAAATAATAAATAATAATAATAATAACTTTAAAAAACAAAAATTAATTAAA 3621
              ||||||   |||| |   || ||| ||||
Db      26735 CCCTGTCTAAAATAATAATAATAATAATAATAATAAAATTAAATTAAAATTAAA 26794

Qy       3622 TT 3623
              ||
Db      26795 AT 26796


RESULT 6
US-10-380-124-10
; Sequence 10, Application US/10380124
; Publication No. US2004005387A1
; GENERAL INFORMATION:
;   APPLICANT: Isis Pharmaceuticals, Inc.
;   APPLICANT: Brett P. Monia
;   APPLICANT: Susan M. Freier
;   TITLE OF INVENTION: ANTISENSE MODULATION OF CLUSTERIN EXPRESSION
;   FILE REFERENCE: RTS-0156
;   CURRENT APPLICATION NUMBER: US/10/380,124
;   CURRENT FILING DATE: 2003-03-10
;   NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
;   LENGTH: 8133
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-380-124-10

Query Match             3.3%;    Score 187.6;  DB 18;  Length 8133;
Best Local Similarity  74.6%;    Pred. No. 9.3e-29;
Matches   249;  Conservative    0;  Mismatches   84;  Indels    1;  Gaps    1;

Qy       3264 TTTAAATTTTTTAAAAAAATAGATGAGTGTGGTGGCTCATGCCTGTAATCCCAACACTTT 3323
              |||||   || |||||||||
Db       1057 TTTAAAGAGGAAAAAAACTGACTGGGAGCTGGTGGCTCCAGCAGTTGAGACCAGCCTGGCCAACACTTT 1116

Qy       3324 GGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCAGTCAGGGCAACAC 3383
              ||  || |||||| ||  ||||   |||| ||||||||||||||||||||
Db       1117 GAGAGGCCAAGGCGGGAGGGATCATTTGAGGCCAGGAGTTTGAGACCAGCCTGCCAAGAT 1176

Qy       3384 AGCAAGACCCCATATCTAAAAAACAAAAACAAAATTAAAAACTTAAAAATTAAAAAACTTA 3443
              |||||||||||||||||||| |||||||||||   | ||||||
Db       1177 AGTGAGACCCTGTCTCTACAAAATAAAAACTTAAAAACTTAAAAAATTAAAAAAATTGGTGTCA 1236

Qy       3444 CACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGT 3503
              ||||||||||| ||||  ||||||| |||||||||||||  |||||||||||
Db       1237 CACCTGTAGTCTCAGCTATGATTGCACCATTGCATTCCGGGAGGCTGAGGCAGGAGGATCACCTGAGCCTGGGAAGTT 1296

Qy       3504 TGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCGTGGGTGACAGTGCAAGAA 3562
              |||||  |||||||  ||||| ||||||||||
Db       1297 GGAGGCTGCAGTGAGCTATGATCGCACCATTGCACCATGCACCAACAGAGTGAGAC 1356

Qy       3563 GCTGTCTCAAAAAATAAATAAATAAATAAAATAAATAC 3596
```

```
Db     1357  |||||||  |||  |||  ||||||||||||||||||
             CCTGTCTCAAATTAAAAAAATAAAATAAAATAAAATAAC  1390

RESULT 7
US-09-984-429-513
; Sequence 513, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: 53 Human Secreted Proteins
;   FILE REFERENCE: PZ018P2
;   CURRENT APPLICATION NUMBER: US/09/984,429
;   CURRENT FILING DATE: 2001-10-30
;   PRIOR APPLICATION NUMBER: 60/244,591
;   PRIOR FILING DATE: 2000-11-01
;   PRIOR APPLICATION NUMBER: 09/288,143
;   PRIOR FILING DATE: 1999-04-08
;   PRIOR APPLICATION NUMBER: PCT/US98/21142
;   PRIOR FILING DATE: 1998-10-08
;   PRIOR APPLICATION NUMBER: 60/061,463
;   PRIOR FILING DATE: 1997-10-09
;   PRIOR APPLICATION NUMBER: 60/061,529
;   PRIOR FILING DATE: 1997-10-09
;   PRIOR APPLICATION NUMBER: 60/071,498
;   PRIOR FILING DATE: 1997-10-09
;   PRIOR APPLICATION NUMBER: 60/061,527
;   PRIOR FILING DATE: 1997-10-09
;   PRIOR APPLICATION NUMBER: 60/061,536
;   PRIOR FILING DATE: 1997-10-09
;   NUMBER OF SEQ ID NOS: 727
;   SOFTWARE: Patentin Ver. 2.0
;   SEQ ID NO 513
;   LENGTH: 18861
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-984-429-513

Query Match        3.3%;   Score 186;  DB 11;  Length 18861;
Best Local Similarity  76.1%;  Pred. No. 3.1e-28;
Matches  242;  Conservative  0;  Mismatches  75;  Indels  1;  Gaps  1;

Qy    3265  TTAAATTTTTAAAAAAATAGATGAGTGTGGCTCATGCCTGTAATCCCAACACTTTG  3324
            |||||| |||  ||||||||||||||||||||||||||||||||||||||||||||
Db    5557  TAAAAGTTGAAAAAAAAATGACTAGGTGCGAGCGGCTCACATGTCTAATCCCAGCGCTTTG  5616

Qy    3325  GGAAGCGGGTCGGGGACGTAGCTTGAGTCCAGTTTGAGACCAGTCAGGCAACACA  3384
            ||||  ||||||||||||||||||||  |||||||||||| ||||||||||||||
Db    5617  GGAGGCCGAGGTGGGAGAATCACTTGAGCTTCAGGAGTTTGAGACCTGAGCAACATA  5676

Qy    3385  GCAAGACCCCCATATCTAAAAAAACAAAACAAAAATTACCTGGGTATGGTTGTGCTC  3444
            ||||||||||||  ||||  |||   ||||||||| | ||||||  |||||
Db    5677  GCAAGACTTCATCTCTACTAAAAGAAAAGTAAAAATTAGCCAGGCGTGGTGTGCAC  5736

Qy    3445  ACCTGTAGTCCAAGCTACACAGGAGGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGTT  3504
            ||||||||||||||||  |||||||||||||||||||| |||||||||||||||||||||
Db    5737  ACCTGTAGTCCCAGCTACTCAGAAAGCTGAGGCAGGAGGATCACTTGAGTCTGGGAGGTT  5796

Qy    3505  GAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAG  3563
            |||||||||||  ||||||||||  ||||||||||  |||||||||||||||
Db    5797  GAGGCTGCAGTGAGTGAGCCATGATCGCACCACTGCACTCCGAGTGACAGAGTACGACT  5856

RESULT 8
US-10-027-632-181221/c
; Sequence 181221, Application US/10027632
; Publication No. US20020198371A1
```

```
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 181221
;   LENGTH: 611
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-181221

Query Match        3.3%;   Score 185.6;  DB 13;  Length 611;
Best Local Similarity  71.6%;  Pred. No. 6.7e-29;
Matches  272;  Conservative  0;  Mismatches  104;  Indels  4;  Gaps  2;

Qy    3251  AGGCTATAAACCTTTTAAATTTTTAAAAAAATAGATGAGTGTGGCTCATGCCTGTA  3310
            |||| |  |||  ||||||| ||||||||||||||||||||||||||||||||||||
Db     410  AGGTTTTCACTGTTTTAAAAAAAGAAAACAAGGGCTGAGCATGGTGGCTCATGCCTGTA  351

Qy    3311  ATCCCAACACTTTGGGAAGCCGGGTCGGGGACGGCCAAGGCAGGTGAGTCCCAGGAGACCA  3370
            |||||||||||||||||||||| | |||||  ||||||||| |||||| ||||| |||||
Db     350  ATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATTGCTTGAGCCCAGGAGTTTGAGACCA  291

Qy    3371  GTCAGGGCAACACAGCAAGACCCCATATCT---AAAAAACAAAACAAAAATTACC  3427
            ||||||| ||||| |||||||  ||||||       |||||||||||||||||||
Db     290  GCCTGAACAACATAGGGAGACTCTGTCTCTCACAAAAAAAAAAAAAAAAAAATGGC  231

Qy    3428  TGGGTATGGTGTGTCCACCTGGTAGTCCAAGCTACACAGGAGGCTGAGGCAGGAAGGATCA  3487
            ||||| |||||| ||||||||||||| |||||  |||||||||||||||||| |||| |||
Db     230  TGGGGTGGTGGTGGTGCGAGCCTGAGTCCCAGCTACTCAGGAGGCTCAGGAGGCAGGAGGATCA  171

Qy    3488  CTTGAGCCCAGGAGTTGAGGCTGCAGTGAGCCGCTGCTACACT-CAGTCTG  3546
            ||||| ||||||||||||||| |||||||| |||||||  |||  |||||||
Db     170  CTTGAGCCTGGAGGTCAGGCTCAAGGCCTGAGCAACATA  111

Qy    3547  GGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAATAAATAACTTTTAAAAAA  3606
            ||||| ||||| |||||||||||||||||  ||||||||| ||||||||||||
Db     110  GGCTACAGAACCAGATCCTGTCTCTCAAAAAAAAAAAAGAAGAAGAGAACAATA  51

Qy    3607  CAAAAATTAATTAAATTTTA  3626
            ||||||||| |||  |||||
Db      50  TAGATATGTTTCCCTTTCTA  31

RESULT 9
US-10-027-632-181221/c
; Sequence 181221, Application US/10027632
; Publication No. US2003020407S5A9
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181221
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-181221

Query Match   3.3%;   Score 185.6; DB 17;   Length 611;
Best Local Similarity   71.6%;   Pred. No. 6.7e-29;
Matches   272; Conservative   0; Mismatches 104;   Indels   4;   Gaps   2;

Qy   3251 AGGCTATAAACCTTTTAAATTTTTAAAAAATATGAGTGTGTGGCTCATGCCTGTA 3310
            ||||||| || ||||||| |||||| | ||||| |||||||||||||||
Db   410 AGGTTTTCACTGTTTTAAAAAAGAAAACAAGGGGCTGAGCATGTGGCTCATGCCTGTA 351

Qy   3311 ATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCA 3370
            |||||||||||||||||||||||||| |||| ||| ||||||||||||||||||||
Db   350 ATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATTGCTTGAGCCCAGGAGTTTGAGACCA 291

Qy   3371 GTCAGGGCAACACAGCAAGACCCCATATCT---AAAAAACAAAACAAAACAAAATTACC 3427
            |||||| ||||||||||||||| ||||||| ||||||||||||||||||||||||||
Db   290 GCCTGAACAACATAGGGAGCTCTGTCTCTACAAAAAAAAAAAAAAAAAAAATGGC 231

Qy   3428 TGGGTATGGTTGTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGCGAGGATCA 3487
            |||||||| |||||||||||||| ||||||||||||||||||||||||| |||||||
Db   230 TGGGGGTGGTGGTGCGCAGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGGATCA 171

Qy   3488 CTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGGCGCTGCTACACT-CAGTCTG 3546
            ||||||||||| ||||||||||||||||||||||||||||||||||| |||| |||||||
Db   170 CTTGAGCCTCGGGAGGTCAAGGCTGCAGTGAGCCGAGATTGCACCACTGCACTCCAGCCTG 111

Qy   3547 GGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAATAAAAATAACTTTTAAAAA 3606
            |||||||||||| ||||| ||||||||||||||||||||||||| || ||||||
Db   110 GGCTACAGAACCAGATCCTGTCTCAAAAAAAAAAAAAGAATGAAAGAGAACAATA 51

Qy   3607 CAAAAATTAATTAAATTTTA 3626
            ||| ||||||| || |||||
Db   50 TAGATATGTTTCCCTTTCTA 31

RESULT 10
US-10-027-632-115440/c
; Sequence 115440, Application US/10027632
; Publication No. US2002019837lAl
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115440
; LENGTH: 3287
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115440

Query Match   3.3%;   Score 185.4; DB 13;   Length 3287;
Best Local Similarity   76.3%;   Pred. No. 1.7e-28;
Matches   254; Conservative   0; Mismatches 76;   Indels   3;   Gaps   2;

Qy   3265 TTAAATTTTTAAAAAAATAGATGAGTGTGTGGCTCATGCCTGTAATCCCAACACTTTG 3324
            ||| ||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db   843 TTTACTTTTTAATAAGGCAGCTGGGTGCAGTCCAGCAGTTTGAGACCAGCACTTTG 784

Qy   3325 GGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCA 3384
            ||| ||||||||||||||| |||||||||||||||||||||||||
Db   783 GGAGGCCAAGGCAGGTGGATTGTTTGAGTTGAGACCAGGAGTTTGAGACCAGCAACATG 724

Qy   3385 GCAAGACCCCATATCTAAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGTCTC 3444
            ||| ||| ||||||||||||||||||| ||||||||| |||||||||||||||
Db   723 GCGAAAACCCATCTCTACAAA--AATACAAAAAAAGTTAGCGGGTGTGTGAGCAGAC 666

Qy   3445 ACCTGTAGTCCAAGTCACACAGAAGCTGAGGCAGAAGGATCACTTGAGCCCAGGAGTT 3504
            |||||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db   665 ACCTGTAGTCCCAGCTACTTAGGAGGCTCAGGTGGGAGGATCACCTGAGCCCAGGAGTT 606

Qy   3505 GAGGCTGCAGTGATCCATGAACGGCGCGTGCTACACT-CAGTCTGGGGTGACAGTGCAAGAAG 3563
            |||||||||||||||||||| |||||||||||| ||| |||||||||||||||||||||||
Db   605 GAAGCTGCAGTGAGCTGAGCTGTGATCGTGCCACTCCAGACTCCAGCTGGGTGATAGAGTGAGACA 546

Qy   3564 CTGTCTCAAAAATAATAAATAAAATAAAAATAAC 3596
            ||||||| |||| || || ||||||||||||
Db   545 CTATCTCAAAAAAAAAAAAAAAAAAAAAAAGAC 513

RESULT 11
US-10-027-632-115440/c
; Sequence 115440, Application US/10027632
; Publication No. US20030204075A9 .
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 115440
; LENGTH: 3287

; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-115440

Query Match   3.3%;   Score 185.4; DB 17; Length 3287;
Best Local Similarity   76.3%;   Pred. No. 1.7e-28;
Matches   254;   Conservative   0; Mismatches   76;   Indels   3;   Gaps   2;

Qy   3265 TTAAATTTTTTAAAAAAATAGATGAGTGTGGCTCATGCCTGTAATCCCAACACTTTG 3324
           ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   843 TTTACTTTTTTAATAAGGCAGCTGGGTGCACAGTTTGAGACCAGCTTTG 784

Qy   3325 GGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGCGGCAACACA 3384
           |||| |||| ||||||||||||||||||||||||||||||||||| ||||||||||
Db   783 GGAGGCCAAGGTGGGTGGATTGTTGTTGAGTTCAGGAGTTTGAGACCAGCCTGGGCAACATG 724

Qy   3385 GCAAGACCCCATATCTAAAAAAACAAAACAAAATTACCTGGGTATGGTTGTGCTC 3444
           ||||||||||| ||||||||||||||||||||| |||||||||||||||||||||
Db   723 GCGAAAACCCATCTCTACAAAA--AATACAAAAAAAGTTAGCCGGGTGTGGTAGCAGAC 666

Qy   3445 ACCTGTAGTCCAAGCTACACAGGAGGCTGAGGCAGGAGAATCATTGAGGCCAGGAGGTT 3504
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   665 ACCTGTAGTCCCAGTTACTTGGAGGCTACGTGGTGGGATCAACTGAGGCCCAGGAGGTT 606

Qy   3505 GAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTCGGGTGACAGTGCAAGAAG 3563
           ||||||||||||| ||||||||||||||||||||| |||| |||||||||||||||||
Db   605 GAAGCTGCAGTGAGCTGTGATCGTGTGCCACTGCACTCCAGACTGGGTGATAGAGTGAGACA 546

Qy   3564 CTGTCTCAAAATATATAAATAAATAAAAATAAC 3596
           ||||||||||||||||||||||| ||||||| ||
Db   545 CTATCTCAAAAAAAAAAAAAAAAAAAAAAAAAGAC 513

RESULT 12
US-10-496-011-4/c
; Sequence 4, Application US/10496011
; Publication No. US2005011858A1
; GENERAL INFORMATION:
;   APPLICANT: BOURGERON, THOMAS
;   APPLICANT: JAMAIN, STEPHANE
;   APPLICANT: QUACH, HELENE
;   APPLICANT: BETANCUR, CATALINA
;   APPLICANT: LEBOYER, MARION
;   APPLICANT: GILLBERG, CHRISTOPHER
;   TITLE OF INVENTION: POLYNUCLEOTIDE AND PROTEIN INVOLVED IN SYNAPTOGENESIS, VARIANTS
;   TITLE OF INVENTION: THEREOF, AND THEIR THERAPEUTIC AND DIAGNOSTIC USES
;   FILE REFERENCE: 253820USOXPCT
;   CURRENT APPLICATION NUMBER: US/10/496,011
;   CURRENT FILING DATE: 2004-05-28
;   PRIOR APPLICATION NUMBER: PCT/FR02/04134
;   PRIOR FILING DATE: 2002-11-28
;   PRIOR APPLICATION NUMBER: CA2364106
;   PRIOR FILING DATE: 2001-11-30
;   NUMBER OF SEQ ID NOS: 98
;   SOFTWARE: PatentIn version 3.3
; SEQ ID NO 4
;   LENGTH: 335199
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (10156)..(10298)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (108500)..(109001)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (205869)..(205928)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (209526)..(209679)
;   FEATURE:
;   NAME/KEY: exon

;   LOCATION: (235028)..(235139)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (238090)..(238212)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (310597)..(310783)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (316139)..(316929)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: (326822)..(330136)
; US-10-496-011-4

Query Match   3.2%;   Score 183; DB 21; Length 335199;
Best Local Similarity   65.2%;   Pred. No. 5.6e-27;
Matches   303;   Conservative   0; Mismatches   155;   Indels   7;   Gaps   2;

Qy   3158 TTTCTGTTCCTCTTAGCTTAGTAACTTTAGGATTTTTAAATAACAACTATTGAAATCATG 3217
           |||||||||||||||||||||||||||||||||||
Db 239974 TTTATATATCTTATATAAAGTCATCTTTTATATATATATAAAATATATTATATATTATATTAATT 239915

Qy   3218 ACATACGTTTAAATGATATTATTTAAAATACGTTAGGCTATAAAACCTTTTAAATTTTTTAA 3277
           |||||||||||||||||||||||||||||||||||
Db 239914 ATGTTACAAAAATGTAATATATATATATAATAAATCATTCAGTGACATTGAAGAGAAAAA 239855

Qy   3278 AAAAATAGATGAGTGTGGCTCATGCCTGTAATCCCAACACTTTGGGAGCCGGGTCG 3337
           ||||||||||||||||||||||||||||||||||||||||
Db 239854 AAAATTGGCACCAGTGCCGGTGTCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTG 239795

Qy   3338 GGAGGATAGCTTGAGTCCAGCAGTTTGAGACCCAGCAGTTTGAGACCGTCGGCAACAGAAGACCCCATA 3397
           |||||||||||||||||||||||||||||||||||
Db 239794 GGTGGATTGGCTCGAGCTCAGCGAGTTTGAGACCAGCCTGGGCAACACAGTGAAACACCCCATC 239735

Qy   3398 TCTAAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGTCCACCTGTAGTCCAA 3457
           |||| ||||| ||||||||||||
Db 239734 TCTATGGAAATACAA------AAAATTAGTTGAGCGTGGTGGTGCACACCTGTAGTCCCA 239681

Qy   3458 GCTACACAGGAAGCTGAGGCAGGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGA 3517
           |||||| ||||||||||||||||||||||||||||||
Db 239680 GCTACTGGGGAGGCTGAGACAGGAGAATTGCTTGAACCCAGGAGGCAGAGGTTGCAGTGA 239621

Qy   3518 TCCATGAACGCGCTGCTACACT-CAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAAT 3576
           ||||||||| || |||||||| |||||||||||||||||||
Db 239620 GCTGAGATCACGCCATTGCACTCCAGCCTGGGTGACAGAGCAAGAACTCTGTCTCAAAAAA 239561

Qy   3577 AATAAATAAAAATAAATAACTTTTAAAAAACAAAATTAATTAAA 3621
           | |||||||||||||||||||||||||
Db 239560 ATAAAATAAAATAAAATAAAATAGAAAAAGAAAAAGAAAGAAACCAAAA 239516

RESULT 13
US-10-322-281-808
; Sequence 808, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;   APPLICANT: David W. Morris
;   APPLICANT: Marc S. Malandro
;   TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;   FILE REFERENCE: 52945201000
;   CURRENT APPLICATION NUMBER: US/10/322,281
;   CURRENT FILING DATE: 2002-12-17
;   NUMBER OF SEQ ID NOS: 866
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 808
;   LENGTH: 47903
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-322-281-808

Query Match   3.2%;   Score 182.6; DB 19; Length 47903;
Best Local Similarity   73.3%;   Pred. No. 2.6e-27;

Matches 247; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

```
Qy    3279 AAAATAGATGAGTGGTGGTCGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGG 3338
           ||||| |||| |||  ||| |||  ||||||||||||||||||||||||||||||||
Db   41168 AAATAGCCAGGTGCTATGTGTTCACGTCTCACAACCCAGCACTTTGGGAAGCTGACGCAG 41227

Qy    3339 GAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGCCAAGACCCCATAT 3398
           |||||||| |||||| ||  |||  ||||||||||| ||||||||||||||  |||||
Db   41228 GAGGATTGCTTGAGCCCAGGAATTTGAGACCAGTCAGTTGACCTGAGACCTCATCT 41287

Qy    3399 CTAAAAAAACAAAACAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAG 3458
           ||||||||||||||||||||||||||||||||||||||  || |||||||| ||||||
Db   41288 CTATTAAAAAAAAAAAAAAAAAAATTAGCCGGGTGGTGCCCACACCTATGGTCCCAG 41347

Qy    3459 CTACACAGGAGCTGAGGCGAAGGGATCACTGAGCCCAGGAGGTTGAGGCTGCAGTGAT 3518
           ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   41348 TCACTTGGGAGGCTGAGGCAGGAGGATCGCTTGAGTCTGAGAGGTTGAACGTGCAATGAG 41407

Qy    3519 CCATGAACGCGCCTGCTCACACT-CAGTCTCGGGTGACAGTGCAAGAGAGCTGTCTCAAAAATA 3577
           ||||| || ||||||| |||||| |||||||||||||| |||  |||||
Db   41408 CCATGTTGGACCACTGCACTCCAGCCTGGGTAACAGAGCAAGACTCTGTTTCAAAAAAT 41467

Qy    3578 ATAAATAAATAAAATAACTTTTAAAAAACAAAAATT 3614
           ||||||||||||||||| ||||||| ||
Db   41468 AAAAATAAATAAATAAAAATTCTAAATAAATAAAACT 41504
```

RESULT 14
US-10-027-632-17813
; Sequence 17813, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;　APPLICANT: Wang, David G.
;　TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;　TITLE OF INVENTION: Polymorphisms in the Human Genome
;　FILE REFERENCE: 108827.129
;　CURRENT APPLICATION NUMBER: US/10/027,632
;　CURRENT FILING DATE: 2002-04-30
;　PRIOR APPLICATION NUMBER: US 60/218,006
;　PRIOR FILING DATE: 2000-07-12
;　PRIOR APPLICATION NUMBER: US 60/198,676
;　PRIOR FILING DATE: 2000-04-20
;　PRIOR APPLICATION NUMBER: US 60/193,483
;　PRIOR FILING DATE: 2000-03-29
;　PRIOR APPLICATION NUMBER: US 60/185,218
;　PRIOR FILING DATE: 2000-02-24
;　PRIOR APPLICATION NUMBER: US 60/167,363
;　PRIOR FILING DATE: 1999-11-23
;　PRIOR APPLICATION NUMBER: US 60/156,358
;　PRIOR FILING DATE: 1999-09-28
;　PRIOR APPLICATION NUMBER: US 60/146,002
;　PRIOR FILING DATE: 1999-08-09
;　NUMBER OF SEQ ID NOS: 325720
;　SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17813
;　LENGTH: 736
;　TYPE: DNA
;　ORGANISM: Human
US-10-027-632-17813

Query Match　　3.2%;　Score 182.4;　DB 13;　Length 736;
Best Local Similarity　68.2%;　Pred. No. 3.5e-28;
Matches 266;　Conservative 1;　Mismatches 122;　Indels 1;　Gaps 1;

```
Qy    3274 TTAAAAAATAGATGAGTGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGG 3333
           | |||||  ||||| ||||| |||||||||||||||||||||||||||||||||||
Db    115 TAAAAATGATAGCCGGGGGGCGTGGGGCTCATACCTGTAATCCCAGCACTTTGGGAGGCTGA 174

Qy    3334 GTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGCCAAGACCC 3393
           |||||||||||||  |||||| || ||  ||||||||||  ||||||||||||||||
Db    175 GGTGGGTGGATCACTTGAGGCCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCC 234

Qy    3394 CATATCTAAAAAAACAAAACAAAACAAAAT
```

RESULT 15
US-10-027-632-17813
; Sequence 17813, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;　APPLICANT: Wang, David G.
;　TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;　TITLE OF INVENTION: Polymorphisms in the Human Genome
;　FILE REFERENCE: 108827.129
;　CURRENT APPLICATION NUMBER: US/10/027,632
;　CURRENT FILING DATE: 2002-04-30
;　PRIOR APPLICATION NUMBER: US 60/218,006
;　PRIOR FILING DATE: 2000-07-12
;　PRIOR APPLICATION NUMBER: US 60/198,676
;　PRIOR FILING DATE: 2000-04-20
;　PRIOR APPLICATION NUMBER: US 60/193,483
;　PRIOR FILING DATE: 2000-03-29
;　PRIOR APPLICATION NUMBER: US 60/185,218
;　PRIOR FILING DATE: 2000-02-24
;　PRIOR APPLICATION NUMBER: US 60/167,363
;　PRIOR FILING DATE: 1999-11-23
;　PRIOR APPLICATION NUMBER: US 60/156,358
;　PRIOR FILING DATE: 1999-09-28
;　PRIOR APPLICATION NUMBER: US 60/146,002
;　PRIOR FILING DATE: 1999-08-09
;　NUMBER OF SEQ ID NOS: 325720
;　SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17813
;　LENGTH: 736
;　TYPE: DNA
;　ORGANISM: Human
US-10-027-632-17813

Query Match　　3.2%;　Score 182.4;　DB 17;　Length 736;
Best Local Similarity　68.2%;　Pred. No. 3.5e-28;
Matches 266;　Conservative 1;　Mismatches 122;　Indels 1;　Gaps 1;

```
Qy    3274 TTAAAAAATAGATGAGTGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGG 3333
           | |||||  ||||| ||||| |||||||||||||||||||||||||||||||||||
Db    115 TAAAAATGATAGCCGGGGGGCGTGGGGCTCATACCTGTAATCCCAGCACTTTGGGAGGCTGA 174

Qy    3334 GTCGGGAGGATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGCCAAGACCC 3393
           |||||||||||||  |||||| || ||  ||||||||||  ||||||||||||||||
Db    175 GGTGGGTGGATCACTTGAGGCCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCC 234

Qy    3394 CATATCTAAAAAAACAAAACAAAACAAAAT 3453
           ||| || ||||||||||||||| |||||||
Db    235 CATCTCTACTAAAAAAAATAAAATTAGCCGGGCGTGGTGGCGGGCGCCTGTAGTCCCAGCTACT 294

Qy    3454 CCAAGCTACACAGGAGCTGAGGCAGGAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCA 3513
           |||||| ||| || || |||||||| |||  |||| |||  |||||| | || |||||
Db    295 TCCAACTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAAGAGGCGGAGGTTGCA 354

Qy    3514 GTGATCCATGAACGCGCCTGCTCACAC-TCAGTCTCGGGTGACAGTGCAAGAGAAGCTGTCTCAA 3572
           ||||||||| || ||  |||||||| | |||  ||||||||| ||  |||||||| ||||||
Db    355 GTGAGCCGAAATCGAGCCACTGCACTTCAGCCTGGGTGACAGACAAGAGACTCGTCTCAA 414

Qy    3573 AAATAATAAATAAATAACTTTTAAAAAATTTAAATTTAAAAACA 3632
           |||||||||| ||||| |||||| |||  ||  ||  |||||||
Db    415 AAAAAAAAAAAAAAAAAATGAAAATGATAGTCTCAACCTCTTGGGATATGGGGAGAGA 474

Qy    3633 CAACACACTAGAGATGTTTGCAAATTGATT 3662
           ||  | | ||| || ||| | |||
Db    475 TAAGTGAAAAGTAATGTCTGCAGAGCGCTT 504
```

```
Qy   3514 GTGATCCATGAACGCGCTGCTACAC-TCAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAA 3572
          ||||| || ||| ||| |||| ||||  |||||||||| ||||||||| |||||||||||
Db    355 GTGAGCCGAAATCGAGCCACTTGCACTTCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAA 414

Qy   3573 AAATAATAATAAATAAATAACTTTTAAAAACAAAAATTAATTAAATTTAAAAACA 3632
          ||||| ||| ||| |||||  ||| |   | ||||| |   || || | |||||||
Db    415 AAAAAAAAAAAAAAAATGAAAATGATAGTCTCAACCTCTTGGGATATGGGGAGAGA 474

Qy   3633 CAACACACTAGAGATGTTTGCAAATTGATT 3662
          || | |||| ||| ||||||||||| ||||
Db    475 TAAGTGAAAAGTAAATGTCTGCAGAGCGCTT 504
```

Search completed: June 23, 2005, 15:00:51
Job time : 3074 secs

```
          GenCore version 5.1.6
     Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       June 22, 2005, 14:47:40 ; Search time 16574 Seconds
                            (without alignments)
                         13035.646 Million cell updates/sec

Title:        US-10-030-294-1
Perfect score: 5676
Sequence:     1 cagctgttcagggatgctg........tttatataattgttccgt 5676

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:     34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters:     68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	427.8	7.5	523	8	AQ168656	AQ168656 HS_3165_B
2	293.2	5.2	638	6	BY723946	BY723946 RST31802
3	280.4	4.9	293	4	BG212218	BG212218 RST31802
4	267	4.7	288	4	BG219090	BG219090 RST38843
5	262	4.6	277	4	BG212753	BG212753 RST32348
6	184.6	3.3	567	4	BI497128	BI497128 df132c05.
7	181.6	3.2	664	8	AQ275866	AQ275866 CITBI-E1-
8	180.6	3.2	725	8	CC061176	CC061176 MUGQ_CH25
9	180	3.2	828	8	BZ601705	BZ601705 WHABF61TR
10	179.8	3.2	733	8	AQ052879	AQ052879 RPCI11-42
11	179.6	3.2	567	8	BQ691256	BQ691256 AGENCOURT
12	179.6	3.2	810	8	AQ780966	AQ780966 HS_3138_B
13	179	3.2	637	8	BQ688837	BQ688837 AGENCOURT
14	178.8	3.2	440	5	AI362915	AI362915 qy81c06.x
15	178.2	3.1	674	9	AG073447	AG073447 Pan trogl
16	178	3.1	463	5	BQ883640	BQ883640 AGENCOURT
17	177.8	3.1	449	1	AI365624	AI365624 ap20e10.x
18	177.6	3.1	680	8	AQ750933	AQ750933 HS_5574_B
19	177.6	3.1	1025	5	BM907570	BM907570 AGENCOURT
20	176	3.1	680	8	AQ046675	AQ046675 RPCI11-42
21	176	3.1	706	4	BM701690	BM701690 UI-E-CQ1-
22	175.4	3.1	330	1	AA226173	AA226173 nc11g08.s
23	175.4	3.1	626	5	BX952768	BX952768 DKFZp781E
24	175.4	3.1	1710	3	CR590862	CR590862 full-leng

```
c 25  175    3.1  175    5  BU153359  BU153359  AGENCOURT
  26  174.2  3.1  174.2  8  AQ343025  AQ343025  RPCI11-12
  27  174.2  3.1  174.2  8  AQ749761  AQ749761  HS_5573_A
  28  174    3.1  174    6  CA395712  CA395712  c60a08.y
c 29  173.6  3.1  173.6  1  AI167878  AI167878  o291b12.x
c 30  173.4  3.1  173.4  6  CA307465  CA307465  UI-H-FT1-
c 31  173.2  3.1  173.2  7  CR546471  CR546471  DKFZp470B
  32  173.2  3.1  173.2  3  CR749594  CR749594  Homo sapi
  33  173.2  3.1  173.2  3  BC036681  BC036681  Homo sapi
c 34  172.8  3.0  172.8  8  CC061208  CC061208  MUGQ_CH25
  35  172.4  3.0  172.4  8  BE562634  BE562634  601336472
c 36  172.2  3.0  172.2  9  AG087187  AG087187  Pan trogl
  37  172    3.0  172    7  W07317    W07317    za94b10.r1
c 38  171.8  3.0  171.8  7  AA715814  AA715814  nw25c01.s
  39  171.8  3.0  171.8  7  CN345184  CN345184  170005999
  40  171.8  3.0  171.8  3  AL831843  HSM803160 Homo sapi
  41  171.6  3.0  171.6  3  AQ413740  AQ413740  RPCI-11-2
  42  171.6  3.0  171.6  5  CR545302  CR545302  DKFZp459I
  43  171.6  3.0  171.6  8  AQ038047  AQ038047  CIT-HSP-2
c 44  171.6  3.0  171.6  8  AL543301  AL543301  AL543301
  45  171.4  3.0  171.4  4  BG536633  BG536633  602566240
```

```
                        ALIGNMENTS

RESULT 1
AQ168656
LOCUS       AQ168656              523 bp    DNA      linear   GSS 16-OCT-1998
DEFINITION  HS_3165_B2_F08_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3165 Col=16 Row=L, genomic survey
            sequence.
ACCESSION   AQ168656
VERSION     AQ168656.1  GI:3566331
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3165 row: L column: 16
            Class: BAC ends
            High quality sequence stop: 523.
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3165 Col=16 Row=L"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Query Match           7.5%;     Score 427.8; DB 8;   Length 523;
Best Local Similarity 90.6%;    Pred. No. 1.4e-59;
```

```
Matches  453;  Conservative    0;  Mismatches   47;  Indels     0;  Gaps      0;

Qy    672 GATCTCATGTCACAGGCTGCCAGAGCTCACGGAACATGCAAGATCATCTGCTTAAGC 731
              ||||  ||    |||||||||||||||||||||| ||||||||||||| |||||
Db     12 GANTCTNTAAGTACGNGATGAATCGATCTACGGAACATGCAAGATCATCTGCTNTAAGC 71

Qy    732 CTCTTGTGGTGGCATCTGTTGTTTCCACTGCCCTGTACCTATTGCTCTTTCCTTGGTTA 791
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 CTCTTGTGGTGGCATCTGTTGTTTCCACTGCCCTGTACCTATTGCTCTTTCCTTGGTTA 131

Qy    792 ACAGAACCTTTATTTTCTCTGAAAACTCTCTGCTCAGTCATGGTAGGGCCATCAGTCCA 851
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 ACAGAACCTTTATTTTCTTCTGAAAACTCTCTGCTCAGTCATGGTAGGGCCATCAGTCCA 191

Qy    852 CATGATCAGGCCTCCTCCTGGCCAAACATGGCATCTTCTTTTGGGAATTTGAATCTTAAG 911
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 CATGATCAGGCCTCCTCCTGGCCAAACATGGCATCTTCTTTTGGGAATTTGAATCTTAAG 251

Qy    912 CTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTACAGTGGCTTTGC 971
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 CTGAATAGCTGAAGTTCAAAAAAAGCTGTTGAATCTGACTTACGCCTACAGTGGCTTTGC 311

Qy    972 AAAGTGACTGTCCATTCCATTCCTTAAGTCCCTGAATTTATAATTTATCCTGGTTACAG 1031
              ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    312 AAAGTGACTGTCCATTCCTATTCCTTAAGTCCCTGAATTTATAAATTTATCCTGGGTACAG 371

Qy   1032 CCCTTTCTGAGATGTGTGGTTTTTTTTCCAACTGTCTCTTATAGTCTGTGAATTTTCATA 1091
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    372 CCCTTTCTGAGATGTGTGGATGTTTTTCCAACTGTCTCTTATAGTCTGAGAATTATCATA 431

Qy   1092 TTTCTTTTCATACATTTTCATGTTTTGTTTGTTTGTGTTTTTTGGCTTTAGGTA 1151
              ||||||||||||| ||||||  ||||||||||| |||  |||  |||||
Db    432 TTTCTTTTCATACATTTCTCATGCTGTGTTTGTTTGTTTGTCCGTGATGNTGTCAGACTTAAGTA 491

Qy   1152 GGCAGAATCAGTTCTGTTG 1171
              |||||   |||| |  |||
Db    492 GAAAGAATGGATCTATTCTG 511


RESULT 2
BY723946
BY723946 RIKEN full-length enriched, adult male hypothalamus Mus
LOCUS      BY723946                638 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION BY723946 RIKEN full-length enriched, adult male hypothalamus Mus
           musculus cDNA clone A230061P11 5', mRNA sequence.
ACCESSION  BY723946
VERSION    BY723946.1  GI:27137063
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1  (bases 1 to 638)
AUTHORS    Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
           Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
           Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
           Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
           Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
           Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
           Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
           Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
           Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
           Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
           Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
           Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
           Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
           Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
           Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
           Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
           Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
           Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
           Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
           Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
           Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
           Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
           Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
           Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
           Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
           Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
           Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE      Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
MEDLINE    22254683
PUBMED     12466851
COMMENT    Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
           Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
           Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
           Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
           Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
           Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
           Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
           Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
           Direct Submission
           Computational Analysis of Full-Length Mouse cDNAs Compared with
           Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
           Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new
           genes. Genome Res. 10 (10), 1617-1630 (2000)
           RIKEN integrated sequence analysis (RISA) system-384-format
           sequencing pipeline with 384 multicapillary sequencer. Genome Res.
           10 (11), 1757-1771 (2000)
           Computer-based methods for the mouse full-length cDNA
           encyclopedia: real-time sequence clustering for construction of a
           nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
           RIKEN Mouse Gene
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site (http://genome.gsc.riken.go.jp) for
           further details.
FEATURES             Location/Qualifiers
     source          1..638
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A230061P11"
                     /sex="male"
                     /tissue_type="hypothalamus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hypothalamus"
                     /note="Site 1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAATTAATCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
```

ORIGIN

FEATURES   Location/Qualifiers
      source   1..293
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."

ORIGIN

Query Match   4.9%;   Score 280.4; DB 4; Length 293;
Best Local Similarity 99.6%;   Pred. No. 1.6e-35;
Matches 281; Conservative   0; Mismatches 1; Indels   0; Gaps   0;

Qy   3117 TCTTTTCATAATTACTTATTTAAAATTGTTGCCATTTAATTCTGTTCCTCTTAGCTT 3176
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   12 TTTTTTCATAATTACTTATTTAAAATTGTTGCCATTTAATTCTGTTCCTCTTAGCTT 71

Qy   3177 AGTAACTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 3236
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72 AGTAACTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 131

Qy   3237 TATTTAAATACGTTAGGCTATAAACCTTTTAAAAATTTTTAAAAAATAGAGTAGTGTGGT 3296
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 TATTTAAATACGTTAGGCTATAAACCTTTTAAAAATTTTTAAAAAATAGAGTAGTGTGT 191

Qy   3297 GGCTCATGCCGTAATCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCA 3356
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 GGCTCATGCCGTAATCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCA 251

Qy   3357 GCAGTTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATAT 3398
             ||||||||||||||||||||||||||||||||||||||||||
Db   252 GCAGTTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATAT 293

RESULT 4
BG219090/c
LOCUS   BG219090   288 bp   mRNA   linear   EST 21-APR-2001
DEFINITION RST38843 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG219090
VERSION   BG219090.1 GI:13745111
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 288)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
  TITLE   Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
   MEDLINE 21227151
   PUBMED   11329013
  COMMENT   Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 286.
FEATURES   Location/Qualifiers
      source   1..288
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

ORIGIN

                     FLC I."

Query Match   5.2%;   Score 293.2; DB 6; Length 638;
Best Local Similarity 77.6%;   Pred. No. 1.1e-37;
Matches 399; Conservative   0; Mismatches 98; Indels   17; Gaps   3;

Qy   5163 AAAACATTTTAGCCCCAGAAATAGTCACAGAAATCCTCAAATCAAACCAGTATCCAGATA 5222
             |||||||||||||||||||||||||||| ||||| ||||| | |||||||
Db   1 AAAACATTGTAGCCACCAGCAGAACTGACAGGAGCTCTCAAATCAAGTCAGAATACAGATA 60

Qy   5223 CAAGGAAGTGTTATGTAGCTGGAGCAGGGTGGACACTCATCAGTTCAGTTACAAA 5282
             ||||||||| ||| |||||||||| || ||||| ||| ||||| |
Db   61 CAAGGAGATGTTATTCAGTTGGAGCAAGGGGACATTTATTAGCTTCAGT------GACA 113

Qy   5283 AGTCCAGGCTGCTGAAATAAACTCTGATGCCATTCATGCCCAGCATCCAATCACGACAGA 5342
             ||||| |||| |||| ||||||| |||||||||||| |||||| | || ||
Db   114 AGTCCTGGCTTCTGTGATTAAACTCTGATGCCATTCATACCAGCACCCAAGCAA 173

Qy   5343 GATCAGAAGTTCAGAGATGCCTCCAGCTCCAAATTGCCAACAACAAGTGTGGCTACTATA 5402
             ||||||||||||||||||||||||||| ||||||||||||||| || ||||
Db   174 GATCAGAAGTTCAGAGATGCCTACAAAT--------TGCCAACAAGTGTGGCCACTCTA 224

Qy   5403 CGTCAAGGACTCTGAGACCGTGAGAGAGGGGAAGAACAACAGTAGAGAGATGCCCAGC 5462
             |||||||| |||| ||| |||||| | ||||| | || ||| |||||||||
Db   225 CGTCAAGGGCTCTAAAACTGTGGCAGAGAG-GAAGAACAGCTTTACAGGGGGTGCCCAGC 283

Qy   5463 TGGGTAAGAATCAGAGTGTTTATGAAGTTTTAGTCAATTGATGAATCTCATTGGCTAAAATC 5522
             |||||||||||| ||||||||||| |||| | ||||| ||||| ||||||||
Db   284 TGGGTAAGAATTGACGGTTTATGATGTGTCTCGTTACTTGAAGACTCTCATTGGCTGAAAGG 343

Qy   5523 AAGAAACGCTTCCGCCTCTTTGCAAATATGTGAAGGAGAGAAGTGCCTAAACTTCTATG 5582
             |||||| || |||| ||||| ||||||| | ||||| |||| ||||| |||||||| ||
Db   344 AAGAAACGCCCCGCCTCTTTGCAAATCGAGTAAAGGGGGAAGTGTCTAAACTTCTATG 403

Qy   5583 TCTGTAGCATTGACCCTATTGCTTTAGCCTCCCGGCTTTATATCTTATATATACACAG 5642
             |||||| ||| ||||||||||||||||| ||||| | |||| | |||||||||||||
Db   404 TCTGATGGCATTGACCCCTATTGCTTTCAGCCTCCGGCTACATCCTAGATAATTCTCAG 463

Qy   5643 GTATTTGTGTATATTTTTATATAAATTGTTCTCCGT 5676
             ||||| || ||||||||||| |||||||||||||
Db   464 GTGTATATGTATATTTTTATGAATTGCTTCCCAT 497

RESULT 3
BG212218
LOCUS   BG212218   293 bp   mRNA   linear   EST 21-APR-2001
DEFINITION RST31802 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG212218
VERSION   BG212218.1 GI:13733905
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 293)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
  TITLE   Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
   MEDLINE 21227151
   PUBMED   11329013
  COMMENT   Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 293.

/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

ORIGIN

Query Match   4.6%;   Score 262; DB 4;   Length 277;
Best Local Similarity 98.1%; Pred. No. 1.6e-32;
Matches 265; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3117 TCTTTTCATATTAATTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTT 3176
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270 TTTTTCATAAATTATTACTTATTTAAAATTTGTTGCCATTTAATTTCTGTTCCTCTTAGCTT 211

Qy 3177 AGTAACTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 3236
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 AGTAACTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 151

Qy 3237 TATTTAAATACGTTAGGCTATAAAACCTTTAAATTTTTAAAAAAATAGATGAGTGTGGT 3296
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150 TATTTAAATACGTTAGGCTATAAAACCTTTAAATTTTTAAAAAAATAGATGAGTGTGGT 91

Qy 3297 GGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCA 3356
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 GGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCA 31

Qy 3357 GCAGTTTGAGACCAGCCAGTCAGGCGAACACAGC 3386
         ||||||||||||||||||||||||||||||||||
Db   30 GCAGTTTGAGACCAGCCAGTCAGGCGAACACAGC 1

RESULT 6
BI497128
LOCUS   BI497128   567 bp   mRNA   linear   EST 28-AUG-2001
DEFINITION df132c05.y1 Morton Fetal Cochlea Homo sapiens cDNA clone
            IMAGE:2537481 5', mRNA sequence.
ACCESSION   BI497128
VERSION   BI497128.1 GI:15336472
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 567)
AUTHORS   Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
            and Morton,C.C.
TITLE   Isolation of novel and known genes from a human fetal cochlear cDNA
            library using subtractive hybridization and differential screening
JOURNAL   Genomics 23, 42-50 (1994)
MEDLINE   95130111
PUBMED   7829101
COMMENT   Contact: Morton, C. C.
            Departments of Pathology and Obstetrics, Gynecology and
            Reproductive Biology
            Brigham and Women's Hospital
            75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
            Tel: 617 732 7980
            Fax: 617 738 6996
            Email: ccmorton@bics.bwh.harvard.edu
            DNA sequencing and analyses were performed by National Institutes
            of Health Intramural Sequencing Center (NISC; see
            http://www.nisc.nih.gov).
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Plate: LLAM6320 row: F column: 10
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES   Location/Qualifiers
     source   1..567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2537481"

/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

ORIGIN

Query Match   4.7%;   Score 267; DB 4;   Length 288;
Best Local Similarity 98.2%; Pred. No. 2.4e-33;
Matches 270; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3119 TTTTTCATATTAATTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAG 3178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   287 TTTTTTTTTTTACTTATTTAAAATTGTTGCCATTTAATTTCTGTTCCTCTTAGCTTAG 228

Qy 3179 TAACTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 3238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   227 TAACTTAGGATTTTTAAATAACAACTATTGAAATCATGACATACGTTTAAATGATAT 168

Qy 3239 TTTAAATACGTTAGGCTATAAAACCTTTAAATTTTTTAAAAAAATAGATGAGTGTGGTG 3298
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   167 TTTAAATACGTTAGGCTATAAAACCTTTAAATTTTTTAAAAAAATAGATGAGTGTGGTG 108

Qy 3299 CTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGC 3358
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   107 CTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGC 48

Qy 3359 AGTTTGAGACCAGCCAGTCAGGGCAACACACAGCCC 3393
         |||||||||||||||||||||||||||||||
Db   47 AGTTTGAGACCAGCCAGTCAGGGCAACACACAGACCC 13

RESULT 5
BG212753/c
LOCUS   BG212753   277 bp   mRNA   linear   EST 21-APR-2001
DEFINITION RST32348 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG212753
VERSION   BG212753.1 GI:13734440
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 277)
AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
TITLE   Creation of genome-wide protein expression libraries using random
            activation of gene expression
JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE   21227151
PUBMED   11329013
COMMENT   Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 227.
FEATURES   Location/Qualifiers
     source   1..277
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES   Location/Qualifiers
source   1..664
     /organism="Homo sapiens"
     /mol_type="genomic DNA"
     /db_xref="taxon:9606"
     /clone="2517A7"
     /sex="male"
     /cell_type="sperm"
     /clone_lib="CITBI-E1"
     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
     CalTech Human BAC Library D"

ORIGIN

Query Match   3.2%;   Score 181.6; DB 8; Length 664;
Best Local Similarity 72.2%; Pred. No. 1.6e-19;
Matches 249; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

Qy 3283 TAGATGAGTGTGGTGCTCATGCCTGTAATCCCAACACTTTGGGAAGGCCGGGTCGGGAGG 3342
        |||||||| ||||||| || |||||||||||||||||| ||||||||||||| ||||||
Db 639 TGGCTGAGTGTGGTGCTCACACCTGTAATCCCAGCACTTTGGNAGGCTCATGTGGAAGG 580

Qy 3343 ATAGCTTGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGCAACGACCCCATATCTAA 3402
        | ||||||||||||||||||||||||||| ||||| || |||||||||||||||||
Db 579 ACAGCTTGAGCCCAGGAGTTTGAGACTAGCTTGGGCAACACAGGAAAACCTCATCTCTAC 520

Qy 3403 AAAAACAAAAACAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGGTAGTCCAAGCTAC 3462
        |||| ||||||| ||| |||||||||||||||||||||||||||| ||||||
Db 519 AAAAAAAAATTTAAGAAAGTAAACGGGAGTGGTGACGAACACCCATAGTCCCAGCTAC 460

Qy 3463 ACAGGAAGCTGAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCAT 3522
        |||||| |||| ||||| |||||||||||||| |||||||||
Db 459 TCAGGAGGCTGAGGCAGGAGGATGACATGAGCCCAAAAGGTTAAGGCTGCAGTGAGCCGT 400

Qy 3523 GAACGCGCCTGCTACACT-CAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAA 3581
        |||||| |||||| || ||| ||||||||| ||||||||||
Db 399 GATCATGCCACTGCACTCCAGCTGGGCAACAGAGCAAGACCCTGCCTCAAAAGAAAA 340

Qy 3582 ATAAATAAAAATAACTTTTAAAAACAAAAATTAATTAAATTTTA 3626
        || ||||||| |||||
Db 339 GAAAATCAATTTGCTAGTTTCCCATCACCTACAGAATAAAGTTTA 295

RESULT 8
CC061176
LOCUS   CC061176   725 bp   DNA   linear   GSS 28-APR-2003
DEFINITION MUGQ_CH252P002Q4T7_H05_CD314_036 CHORI-252 Vervet Monkey Library
     Cercopithecus aethiops genomic clone CH252-2B10, genomic survey
     sequence.
ACCESSION   CC061176
VERSION   CC061176.1 GI:29790439
KEYWORDS   GSS.
SOURCE   Cercopithecus aethiops (African green monkey)
ORGANISM   Cercopithecus aethiops
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
     Cercopithecinae; Cercopithecus.
REFERENCE   1 (bases 1 to 725)
AUTHORS   Minhas,R., Zhang,X., Dore,C., Villeneuve,A., Lepage,P.,
     Forgetta,V., McKee,K., Ophoff,R.A., Fairbanks,L.A., Freimer,N.B.,
     Ervin,F.R., Palmour,R.M., Hudson,T.J. and Dewar,K.
TITLE   UCLA/MUGQ/St-Kitts Vervet Monkey Mapping Project
JOURNAL   Unpublished (2003)
COMMENT   Other_GSSs: MUGQ_CH252P002Q4SP6_H05_CD315_036
     Contact: Dewar K

/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI;
Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'."

ORIGIN

Query Match   3.3%;   Score 184.6; DB 4; Length 567;
Best Local Similarity 68.6%; Pred. No. 5.5e-20;
Matches 300; Conservative 0; Mismatches 129; Indels 8; Gaps 3;

Qy 3170 TTAGCTTAGTAACTTTAGGATTTTAAATAACAACTATTGAAATCATGACATACGTTTAA 3229
        ||| ||| || |||||| |||||||| || ||||
Db 59 TTGCTTTATAAGTTTTCTCAATATTGTGAAAAACTTCCACATCAGTAGATACATGTCC 118

Qy 3230 ATGATATTATTTAAATACGTTAGGCTATAAACCTTTAAATTTTTAAAAAATAGATGA 3289
        ||||| |||||||| |||||||| ||||||||| |||||||||||
Db 119 ATAATTTTTTTTTGTATTGTTTGATTTTTA---TTAAATTTATTGAAATATAGGCTGG 175

Qy 3290 GTGTGTGGCTCATGCCTGTAATCCCAACACTTTGGGAAGCGCGGGTCGGGAGGATAGCTT 3349
        ||||||||| ||| ||||||||||||||| |||||
Db 176 GTGCAGTGGCTTACCTCTGTAATCCCAACACTTTGAGAGGCTGAGGTGGGATGATTCCTT 235

Qy 3350 GAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGACCCCATATCTAAAAAAAAACA 3409
        ||| ||||||||||||||||||||| |||||||||
Db 236 GAGCTTCAGGAGTTCAAGACCACCCTGGGCAACATATCTCTACAAAAAA 295

Qy 3410 AAACAAAACAAAAATTACCTGGGTATGGTTGTGCTCACCTGGTAGTCCAAGCTACACAGGAA 3469
        ||| |||||||| ||||| |||||||||||||||||||
Db 296 ATAGA---AAAATTAGCTGGGCATGGTGGCGTGCCTGTAGTCCCAGCTACTCGGGAG 351

Qy 3470 GCTGAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGAACGCG 3529
        |||||||| |||||||||||| |||||||||
Db 352 GCTGAGGCGGGGGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCTAGGATTGCA 411

Qy 3530 CTGCTACACT-CAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAATAAATA 3588
        ||||| |||||||||||||| |||||
Db 412 CTGCTGCACTCCAGCCTGGGTGACAGACGAGACCCTGTCTCAAAAACCCAAAAACAA 471

Qy 3589 AAAATAACTTTTAAAAA 3605
        || |||| ||||
Db 472 AACTGAAATATCAAATA 488

RESULT 7
AQ275866/c
LOCUS   AQ275866   664 bp   DNA   linear   GSS 22-NOV-1998
DEFINITION CITBI-E1-2517A7.TF CITBI-E1 Homo sapiens genomic clone 2517A7,
     genomic survey sequence.
ACCESSION   AQ275866
VERSION   AQ275866.1 GI:3901682
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 664)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
     Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
     Venter,J.C.
TITLE   Use of a random human BAC End Sequence Database for Sequence-Ready
     Map Building
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CITBI-E1-2517A7.TR
     Contact: Mark Adams
     Department of Eukaryotic Genomics
     The Institute for Genomic Research

```
DEFINITION  WHABF61TR Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-8K2, genomic survey sequence.
ACCESSION   BZ601705
VERSION     BZ601705.1  GI:31510167
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     End-sequence profiling: Sequence-based analysis of aberrant genomes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
  MEDLINE   22709111
   PUBMED   12788976
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-8K2"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
ORIGIN

Query Match      3.2%;      Score 180;  DB 8;   Length 828;
Best Local Similarity 67.7%;   Pred. No. 2.9e-19;
  Matches 283;   Conservative 0;   Mismatches 130;   Indels 5;   Gaps 2;

Qy    3295 GTGGGCTCATGCCTGTAATCCCAACACTTTGGGAGGCCGGGGTCGGGAGGATAGCTTGAGTC 3354
                |||| |||| |||| ||||||||||||||||||||||| |||||||||| ||||||||||||
Db     401 GTGGTTCATACCTATATATCCCAACACTTTGGGAGGCTGGGGAGGATCACTTGAGAC 460

Qy    3355 CAGCAGTTTGAGACCAGTCAGGGCAACACAGCAAGACCCCATATCTCAAAAAAACAAAACA 3414
                ||||||||||||||||| || ||||||| ||| |||||||||||   |||||||||||||
Db     461 CAGGAGTTTGAGACCAGCAGCCTGGCAACATAGCGAGACCCCATCT---ACATAAAAAAAA 516

Qy    3415 AAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGA 3474
                ||| ||| ||| |||| |||| |||||| ||||||||||| ||||| ||||| ||||||
Db     517 TAAAAATGTTAGCTAGGCATAGTGATATGCACCTGTAGTCCTGTAGTACTGGGAGGCTGA 576

Qy    3475 GGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGGTTGAGGCTGCAGTGAGCGTGCTGCT 3534
                |||||||||||| ||| |||| ||||||| ||||| |||||| ||||| ||| ||||||
Db     577 GGCAGGAGTATCGCTTGCTTGAGACCAGGAGTTTGAGGGTGCGTGAGATGTGCCACT 636

Qy    3535 ACACTC-AGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAAAAT 3593
                 ||||| |||| ||| |||||| ||||| |||| |||||||| ||||| |||||
Db     637 GCACTCCAGCCTGGGAGACAGGAGCAGAGCAAGACAGATCTTGTCTCTCTAAAATAAATAATAAAATAAA 696

Qy    3594 AACTTTTAAAAACAAAATTAATTAAATTTTAAAAATATTTAGAAGAGT 3653
                 ||||||||||| |||| | ||| |||||| ||||| ||| ||||||
Db     697 ATTTTTTTAAAAACAAAGTGCTTGAACTCCTTGAACCAGACCAAGAGACTAAATTTAATCTAA 756

Qy    3654 AAATTGATTATTTGGGAGTCTATATCCCTGAAGTAATATTTAGAAGAGT 3711
                ||| || |||||||| ||| ||||||||||||| || || |||||||||
Db     757 AATTCTTTCTTTTTTGGTATTGATAAATGACAGTTAGTTATGCTGAGTCAAATGACT 814
```

```
McGill University and Genome Quebec Innovation Centre
McGill University
740 Dr. Penfield Room 7214, Montreal, QC, Canada, H3H 1A4
Tel: 514 398 3311 x00089
Fax: 514 398 1795
Email: ken.dewar@mcgill.ca
Plate: 2  row: B  column: 10
Seq primer: T7 : TAATACGACTCACTATAGGG
Class: BAC ends.

FEATURES             Location/Qualifiers
     source          1..725
                     /organism="Cercopithecus aethiops"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9534"
                     /clone="CH252-2B10"
                     /sex="male"
                     /cell_type="White blood cell"
                     /dev_stage="Adult"
                     /clone_lib="CHORI-252 Vervet Monkey Library"
                     /note="Vector: pTARBAC2.1; Site 1: EcoR1; Site 2: EcoR1;
                     Constructed by Michael Nefedov In Pieter de Jong's
                     laboratory at BACPAC Resources, Children's Hospital in
                     Oakland Research Institute."

ORIGIN

Query Match      3.2%;      Score 180.6;  DB 8;   Length 725;
Best Local Similarity 60.1%;   Pred. No. 2.3e-19;
  Matches 335;   Conservative 0;   Mismatches 219;   Indels 3;   Gaps 2;

Qy    3078 AACTCTCGCCATTATTATTCCATAATGCAATAGGAGCCAATCTTTTTCATAATTACTTATT 3137
                |||||  |||||||
Db      21 AACACCCTAACTTATTGGCAAGCTTGTTTATATTTCCAAGTATATTATTTTTGTTTGTT 80

Qy    3138 TAAAAATTTGTGTGCCATTTAATTTCTGTTCCTCTTAGCTTAGTAACTTTAGGATTTTAAA 3197
                |||||   ||||
Db      81 TAAACATGCTATTAATTTTTTACAGATTATTTTTCTTTTGGTGTGCTGAGGTGAAAGAACA 140

Qy    3198 TAACAACTATTGAAATCATGACATACGTTTAAATGATATTAATTAAATACGTTAGGCTAT 3257
                ||||
Db     141 TGTTCTGTATTGTTTATGCTTTGTAAAATTTATTGAATCTTTTTGTGACCTGGTGTAAA 200

Qy    3258 AAACCTTTAAATTTTTAAAAAATAGATGAGTGGTCGCCTGCTCATGCCTGTAATCCCAA 3317
                |||||||
Db     201 ACAATGAACAATATTTGTATGATACTGGCTGGGCACGATAGCTCATGCCTGTAATCCCAG 260

Qy    3318 CACTTTGGGAAGCCGGGTCGGGAGGATAGCTTGAGTCCAGCAGTTGAGACCAGTCAGGG 3377
                |||||||
Db     261 CAGTTTGGGAGGCCGAGGCCCAAGGTGGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGC 320

Qy    3378 CAACACAGCAAGACCCCATATCTCAAAAAACAAAATAATAACAAAATTACCTGGGTATGGT 3437
                |||||
Db     321 CAACATGCAAAACCTCATCTCTACTAAAAAAAAACAAAAAAAA--TAGCTGGGGTGTGGT 378

Qy    3438 TGTGCGTCACCTGTAGTCCAAGCTACACAGGAACGTGAGGCCCAGAGAGGATCACTTGAGCCCA 3497
                ||||
Db     379 TGCACCACACTGTAATCCCAGATACTTGGGAGGCTGAGGCAGGAGAATCACTTAAATCTA 438

Qy    3498 GGAGGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACT-CAGTCTGGGTGACAGTG 3556
                ||||
Db     439 GGAGGCAGGAGGTTACAAGGTGGTGCAAGATTGCACCACTGCACCCTTGGTGACACAGAG 498

Qy    3557 CAAGAAGCTGTCTCAAAAATAATAAATAATAAACTTTAAAAAACAAAATTAA 3616
                ||||
Db     499 AGAGACAACTGTCTCAAAAAAAAAATATATATATACACACACAAATACA 558

Qy    3617 TTAAATTTTAAAAAACAC 3633
                |||
Db     559 TACAAATTTTATGATAC 575
```

```
RESULT 9
BZ601705
LOCUS       BZ601705               828 bp    DNA     linear   GSS 08-JUN-2003
```

```
RESULT 10
AQ052879
LOCUS       AQ052879                733 bp    DNA       linear    GSS 20-APR-1999
DEFINITION  RPCI11-42023.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-42023,
            genomic survey sequence.
ACCESSION   AQ052879
VERSION     AQ052879.1  GI:3349834
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 733)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
            Venter,J.C.
TITLE       Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL     Unpublished (1998)
COMMENT     Contact: Mark Adams
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: mdadams@tigr.org
           Clones are derived from the human BAC library RPCI-11. For BAC
           library availability, please contact Pieter de Jong
           (pieter@dejong.med.buffalo.edu). Clones may be purchased from
           BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
           Research Genetics (info@resgen.com). BAC end search page:
           http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..733
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7516102"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-42023"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN
Query Match     3.2%;  Score 179.8; DB 8;  Length 733;
Best Local Similarity  65.9%;  Pred. No. 3.2e-19;
Matches  276;  Conservative    0;  Mismatches  142;  Indels   1;  Gaps   1;

Qy    3196 AATAACAACTATTGAAATCATGACATACGTTTAAATGATATTATTTAAATACGTTAGGCT 3255
           | |||| ||||||||| |||||| || ||||| || ||||| ||  |||| |||||||||
Db     267 ATTAAGAGATCATTAATCCAGCCACTTATTTTCAAATGAGAAAACGCAGAAAG 326

Qy    3256 ATAAACCTTTAAATTTTTAAAAAAAATAGATGAGTCGGTGGTCCGTGTAATCCC 3315
           |||||| ||||| |||| |||||||| ||||||||| | || ||| |||||||||
Db     327 TTAAATGACTTGTTCATGTTCATACATTGGCTGGGCATGGTGCTCACACCTGTAATTCC 386

Qy    3316 AACACTTTGGGAGCCGGTCGGGAGGATAGCTTGAGTCTCCAGCAGTTTGAGACCAGTTCAG 3375
           ||||||||||||| | ||||||||||||||| ||||||| |||||||| ||||||||||
Db     387 AGCACTTTGGGAGGCCCAGGCTGGAGGAACACTTGACCCCAGGAGTTCGAGATCAGCCTG 446

Qy    3376 GGCAACAACAGCAAGACCCCATATCTAAAAAACAAAACAAAACAAAATTACCTGGGTATG 3435
           |||| ||  ||||||||||||| |||| |||| ||| |||| || |||| ||||||||||
Db     447 GACAACATGGTGAAACCCCATCTCTAAAAAAAAAAAAAAAAAAAATTAGCCAGGCTTT 506

Qy    3436 GTTGTGCTCACCTGTAGTCCAAGCTACACAGGAAGCTGAGGCAGAAGGATCACTTGACC 3495
           || | || || ||| |||| |||| |||| |||||||||||| |||||||| |||||||||
Db     507 GCAGCACATGCCTGTGTCCCAGCTACTTGGAAGGCTGAAGTGGGAGGATTGCTTGAGCC 566

Qy    3496 CAGGAGGTTGAGGCTGAGCGGTGCTGAGTGATCCATGCAACCGCGCTGCTACACT-CAGTCTCGGGTGACAG 3554
           |||||||| |||||||| |||| ||||||||| || |||||||| |||||||||||
Db     381 AGGAGGGCTGGAGGCGGCGGAGATGGCTTGAGCCCAGGAGGTTGAGGCTGAGCAGTGAGCCATGA 440
```

```
Db     567 CAGGAGGTCAAGGCTACAGTGAACCGTGATTGTGTACCACTCACCAGCCTGGGTGACAG 626
           |||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Qy    3555 TGCAAGAAGCTGTCTCAAAAATAATAAATAATAAAATAACTTTAAAAAACAAAT 3613
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     627 AGTGAAACCCTGTCTCAAACAACAACAACAAACAAACAAACAAACAGTTCATACATTAAATT 685
```

```
RESULT 11
BQ691256
LOCUS       BQ691256                567 bp    mRNA      linear    EST 15-JUL-2002
DEFINITION  AGENCOURT_8176628 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6252453
            5', mRNA sequence.
ACCESSION   BQ691256
VERSION     BQ691256.1  GI:21816572
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 567)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLICM2398   row: 1   column: 22
           High quality sequence stop: 542.
FEATURES             Location/Qualifiers
     source          1..567
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6252453"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN
Query Match     3.2%;  Score 179.6; DB 5;  Length 567;
Best Local Similarity  73.0%;  Pred. No. 3.6e-19;
Matches  257;  Conservative    0;  Mismatches  88;  Indels   7;  Gaps   2;

Qy    3285 GATGAGTGTGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGAGGAT 3344
           ||||| | | ||||||||||||||||| ||||||||||||||| |||||| |||||||||
Db     207 GCTGGGTACAGTGACTCACATCTGTAGTCCCAGCCACTTTGGGAGGCCGGGCGGGCAGAT 266

Qy    3345 AGCTTGAGTCCAGCAGTTTGAGACCAGTTCAGCCAACGCCAACACAGCAAGAGACCCATATCTAAAA 3404
           |||||||||| |||||| ||||||| ||||||||||||||||||||| |||||||||||||||
Db     267 CACTTGAGCCCAGGAGTTTGAGACCAGCCTGGGCAACACAGCAAAACCCTTTGTCT---- 322

Qy    3405 AAACAAAACAAAACAAAATTACCTGGGTATGGTGGTGTGCTCAGTTGGTGCCTGTAGTCCCAGCTACAC 3464
           ||| || | ||| ||||| ||||| |||||| |||||||| |||||||| |||||||||||||
Db     323 --ACCAAAATATGAAATTTAGCTGGGCATGGTGGCTGCGGGTTGAGCCCAGGAGGTTGAGGCTGAGC 380

Qy    3465 AGGAAGCTGAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGAGCAGTGATCCATGA 3524
           |||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||
Db     381 AGGAGGGCTGGAGGCGGCGGAGATGGCTTGAGCCCAGGAGGTTGAGGCTGAGCAGTGAGCCATGA 440
```

```
Qy  3525 ACGGCGTGCTACA-CTCAGTCTCGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAAT 3583
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   441 TTGCGCACTGCACCCCAGCCTGGGCAACAGAGCAAGACCCTGTCTCAAAAAATATATA    500

Qy  3584 AAATAAAAATAACTTTTAAAAAACAAAATTAATTAAATTTTAAAAACACAA 3635
         |||| |||| ||| |||||||| ||||||||||||
Db   501 TAAATAAAAAAGTTTTAAAAAGAGAAAAATAAACCNNNNAAAAAAAAACAANA 552
```

RESULT 12
AQ780966
```
LOCUS       AQ780966              810 bp    DNA       linear   GSS 02-AUG-1999
DEFINITION  HS_3138_B1_D10_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3138 Col=19 Row=H, genomic survey
            sequence.
ACCESSION   AQ780966
VERSION     AQ780966.1  GI:5583926
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 810)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3138  row: H  column: 19
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 810.
FEATURES             Location/Qualifiers
     source          1..810
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3138 Col=19 Row=H"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN
Query Match            3.2%;   Score 179.6;  DB 8;  Length 810;
Best Local Similarity 69.9%;   Pred. No. 3.3e-19;
Matches  256;  Conservative   0;  Mismatches 109;  Indels   1;  Gaps   1;
```

```
Qy  3261 CCTTTTAAATTTTTTAAAAAAATAGATGAGTGTGTGGCCTCATGCCTGTAATCCCAACAC 3320
         |||| ||| |||||||||||| ||||| ||||||||||||||||||||||
Db   261 CCTTCTCAATTTAAAAAAAAAAAGCCTGGCCTAGTGGCTCATGCCTATAATCCCAGCAC   320

Qy  3321 TTTGGGAGGCCGGGTCGGGAGGATCAGCAGGTCCAGCAGTTGAGACCAGTCAGGGCAA 3380
         |||||||||||||| |||||||||||| |||||||| |||||||||||
Db   321 TTTGGGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTTGAGCCCAGCCTGGCCAA 380

Qy  3381 CACACGAAGACCCCATATCTAAAAAACAAAACAAAAACAAAATTACCTGGGTATGGTTGT 3440
         ||||||| ||| ||||| |||||| ||||||||||||
Db   381 CATGGTGAAACCCTGTCTTTACCAAAAAAATACAAAA-AAAATTAGCTTGGCGTGGTGGC 439
```

```
Qy  3441 GCTCACCTGTAGTCCAAGCTACACAGGAGGCTGAGGCAGGAGAATCACTTGAGCCCAGGA 3500
         ||||||||||| |||||||||||||||||||||||||||||||||| |||||||||
Db   440 ACATGCCTGTAGTCCCAGCTACTCATGAGGCTGAGGCAGGAGAATCACTTGAACCCAGA  499

Qy  3501 GGTTGAGGCTGCAGTGATCCATGAACGCGCTGCTACACTCAGTCTCGGTGACAGTCAAG 3560
         |||||||| |||||||||||||||||||||| ||||||||||||||||
Db   500 GGCGGAGGTTGCAGTGGCCCAGGATTGTGCCACTGCACTCCAGCCTCGGCGACAGAGCAAG 559

Qy  3561 AAGCTGTCTCAAAAATAATAATAAACTTTTAAAAAACAAAAATTAATTAA 3620
         | |||||||||||||||| ||||||||||||||||||||||
Db   560 ATTCTGTCTCAAAAATAAAAAAAATATATATATATATATATATATATATA 619

Qy  3621 ATTTTA 3626
         ||||||
Db   620 CATATA 625
```

RESULT 13
BQ688837
```
LOCUS       BQ688837              637 bp    mRNA       linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8344575 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6247825
            5', mRNA sequence.
ACCESSION   BQ688837
VERSION     BQ688837.1  GI:21814153
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 637)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2386  row: 1  column: 02
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1..637
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6247825"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN
Query Match            3.2%;   Score 179;  DB 5;  Length 637;
Best Local Similarity 75.4%;   Pred. No. 4.4e-19;
Matches  251;  Conservative   0;  Mismatches  75;  Indels   7;  Gaps   2;
```

```
Qy  3285 GATGAGTGTGTGGCTCATGCCTGTAATCCCAACACTTTGGGAGGCCGGGTCGGGAGGAT 3344
         |||||||||||||| ||||||||||||||||||||||||||||||||
Db   311 GCTGGGTACAGTAACTCACATCTGTAGTCCGTAGTCCCAGCCACTTTGGGAGGCCCGAGATGGGCAGAT 370

Qy  3345 AGCTTGAGTCCAGCAGTTGAGACCAGTCAGGGCAACACAGCAACACAGCAAGAGACCCCATATCTAAAA 3404
         |||| ||||||||||||||||||||||||||||
```

```
Qy  3290  GTGTCGGTGGCTCATGCCTGTAATCCCAACACTTTGGGAAGCCGGGTCGGGGATAGCTT  3349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304  GTGTCGGTGGCTCATACCTGTAATCCCAGCACTTTGGGAAGGCGAGTGGGAGGATCGCTT  245

Qy  3350  GAGTCCAGCAGTTTGAGACCAGTCAGGGCAACAGCAACCCATATCTAAAAAAACA  3409
          |||||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db   244  GAGCCCAGGAGTTCAAGACCAGTCTGGGCAACATAGCAAGACTCTCTCTACAA---A  188

Qy  3410  AAACAAACAAAATTACCTGGGTATGGTGGCGCAGCTGTCAAGTCACACAGGAA  3469
          ||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db   187  AAATACAAAAAATTAGCAGGGCGTGGTGGCGTACCGGTTGTCCCAGCTACTCTGGAG  128

Qy  3470  GCTGAGGCAGAAGGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGACGCG  3529
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127  GCTGAGGTGGGAGGATCACTTGAGCCCAGGAATTTGAGGCTGCATTCAGCTATGATTGCA  68

Qy  3530  CTGCTACACT-CAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAATAATA  3588
          ||||||||| ||||||||||||||||||||||||||||||| ||||| |||||| |
Db    67  CCATTGCACTCCTGCCTGGGAGACAGAGTAAGACTCTGTCTCAAAAAAAAAAGAAAA  8

Qy  3589  AAAATA  3594
          |||||
Db     7  AGAAAA  2
```

RESULT 15
AG073447/c
LOCUS   AG073447   674 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-065D24.F, genomic survey sequence.
ACCESSION   AG073447
VERSION   AG073447.1 GI:16625249
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 674)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
              Sequencing: -21M13
            LIBRARY
              Vector   : pKS145
              R.Site 1 : SacI
              R.Site 2 : SacI.
FEATURES   Location/Qualifiers
     source   1..674
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-065D24.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

Query Match   3.1%; Score 178.2; DB 9; Length 674;
Best Local Similarity   75.9%; Pred. No. 5.8e-19;
Matches 233; Conservative   0; Mismatches 73; Indels 1; Gaps 1;

```
Db   371  CACTTGAGCCCAGGAGTTTGAGACCAGCCAGGCTGGGCAACAGCGGAACCGTTTCTCT----  426
          |||||||||||||||||||||||||||||||||||||
Qy  3405  AAACAAACAAAATTACCTGGGTATGGTGTGTGCTCACCTGGTTCCAAGTCCAAGTCACAC  3464
          || ||| |||
Db   427  --ACCAAAAATATGAAATTTAGCTGGGCATGGTGCTGTGTGTCTGTCCCAGTCCCCAGCTGCTC  484

Qy  3465  AGGAAGCTGAGGCAGGAGATCACTTGAGCCCAGGAGGTTGAGGCTGCAGTGATCCATGA  3524
          |||||||||||||||||||||||||||||||||||||||||
Db   485  AGGAGGCTGAGGCGGGAGGATGGCGTGGAGGTTGAGGCCAGGAGGTTGAGGCTGCAGTGA  544

Qy  3525  ACGGCTGCTACA-CTCAGTCTGGGTGACAGTGCAAGAAGCTGTCTCAAAAATAATAAT  3583
          ||||||||||| || |||||||||||||||||||||||||||||||||| ||||||||
Db   545  TTGCGGCCACTGCACCCCAGCCTGGGCAACAGACCCTGTCTCAAAAATATATATAT  604

Qy  3584  AAATAAAATAACTTTTAAAAAACAAAAATTAA  3616
          |||||||| |||||| ||||||
Db   605  AAATAAAAAGTTTTTAAAAAAAAAAAAAAAA  637
```

RESULT 14
AI362915/c
LOCUS   AI362915   440 bp   mRNA   linear   EST 06-JAN-1999
DEFINITION qy81c06.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2018410 3'
            similar to contains Alu repetitive element:, mRNA sequence.
ACCESSION   AI362915
VERSION   AI362915.1 GI:4114536
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 440)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES   Location/Qualifiers
     source   1..440
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2018410"
                     /clone_lib="DH10B"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="NCI_CGAP_Brn25"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
ORIGIN

Query Match   3.2%; Score 178.8; DB 1; Length 440;
Best Local Similarity   78.4%; Pred. No. 5e-19;
Matches 240; Conservative   0; Mismatches 62; Indels 4; Gaps 2;

```
Qy   3289 AGTGTGGTGCTCATGCCCTGTAATCCAACACTTTGGGAGCCGGGTCGGGAGGATAGCT 3348
          |||| |||||||||||||||||||||||||||| ||||||  ||||  |||||||||||| ||
Db    396 AGTGCAGTGGCTCATGCCTGTAATCCCAACATTTTGGTAGGCCAAGGCGGGAGGATCACT 337

Qy   3349 TGAGTCCAGCAGTTTGAGACCAGTCAGGGCAACACAGCCAAGACCCCATATCTAAAAAAAC 3408
          ||||||||| |||||||||||||||| | |||| |||| |||||| |||| ||||| ||||
Db    336 TGAGTCCAGGAGTTTGAGACCAGCCTGGGCAATACAGAAAGATCCTATTTCTACCAAAA- 278

Qy   3409 AAAACAAAACAAAATTACCTGGGTATGGTTGTGCTCACCTGTAGTCCAAGCTACACAGGA 3468
          | |||||| ||||| || ||||| |||| |||| ||| ||||||||| |||| |||||
Db    277 ATAAAAAATAAAAATGAGCCAGGAGTGCTGGTGGCATTCCTGTAGTCCCAGCTACTCAGGA 218

Qy   3469 AGCTGAGGCAGAAGGATCACTTGAGCCCAGGAGTTGAGGCTGCAGTGATCCATGAACGC 3528
          | ||||||| |||||||||||||||||| ||||| ||||||||  || |||  ||||
Db    217 GTCTGAGGTGGGAGGATCACTTGAGCCCGGGATGTCGAGGCTACACTAAGTGGTGATCAT 158

Qy   3529 GCTGCTACACTCAGTCTCGGGTGACAGTGCAAGAAGCTGTCTCAAAAAATAATAAATAAATA 3588
          || |||||| |||||||| |||||||| || | ||||||| ||||||||| | | ||||
Db    157 GCCACTGCACTCAATCTGGGGTGATAGAGTGAGACCCCTGTCTCAAAAAAATTTAAAAAAGA 98

Qy   3589 AAAATAA 3595
          || |||
Db     97 ACTATGA 91
```

Search completed: June 23, 2005, 12:27:16
Job time : 16580 secs